Vol. 52 OCTOBER 1961 No. 5

# Journal of Educational Psychology

## CONTENTS

*Personality Factors among Science and Technology Freshmen* ..... 223
EUGENE B. NADLER AND GILBERT K. KRULEE

*The Draw-A-Man Test in Indonesia* ..... 232
R. MURRAY THOMAS AND ANWAR SJAH

*American Graduate Students: Some Characteristics of Aspiring Teachers and Researchers* ..... 236
DAVID GOTTLIEB

*The Acquisition and Retention of Paired Associates by Good, Average, and Poor Readers* ..... 241
WAYNE OTTO

*The Contribution of Response in Mental Measurement* ..... 249
WILLAVENE WOLF AND JAMES B. STROUD

*Effects of Training and Anxiety upon Teachers' Preferences for Information about Students* ..... 254
PHILIP J. RUNKEL AND DORA E. DAMRIN

*Learning in Fourth and Sixth Graders as a Function of Sensory Mode of Stimulus Presentation and Overt or Covert Practice* ..... 262
JOAN LOCKARD AND JOSEPH B. SIDOWSKI

*The Role of Discriminability in Meaningful Verbal Learning and Retention* ..... 266
DAVID P. AUSUBEL AND DONALD FITZGERALD

AMERICAN PSYCHOLOGICAL ASSOCIATION

# Journal of Educational Psychology

RAYMOND G. KUHLEN, Editor
*Syracuse University*

Advisory Editors

JOHN B. CARROLL
*Harvard University*
LEE J. CRONBACH
*University of Illinois*
FRANCIS J. DIVESTA
*Syracuse University*
CHESTER W. HARRIS
*University of Wisconsin*
ERNEST R. HILGARD
*Stanford University*
NICHOLAS HOBBS
*George Peabody College for Teachers*
DOROTHEA MCCARTHY
*Fordham University*
CARSON MCGUIRE
*University of Texas*
PAUL H. MUSSEN
*University of California, Berkeley*
T. ERNEST NEWLAND
*University of Illinois*
H. H. REMMERS
*Purdue University*
JULIAN C. STANLEY
*University of Wisconsin*
J. B. STROUD
*State University of Iowa*
DONALD E. SUPER
*Teachers College, Columbia University*
GEORGE G. THOMPSON
*Ohio State University*

---

The *Journal of Educational Psychology* publishes original investigations of problems of learning, teaching, and the psychological development of the individual. It contains articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, personality, social relations, and vocational orientation, particularly (but not exclusively) as related to school adjustment or the impact of the school program upon development and adjustment; emotion, motivation, and character; mental development; and methods. The last includes tests, statistical techniques, and research techniques in cross-sectional and developmental studies. Studies published deal with all levels of education—preschool, elementary, secondary, college, and adult education—and with all age groups.

Manuscripts and correspondence on editorial matters should be addressed to the Editor: Raymond G. Kuhlen, Department of Psychology, Syracuse University, 123 College Place, Syracuse 10, New York.

All manuscripts should be submitted *in duplicate* to facilitate referral to advisory editors and must be prepared to conform to the style and procedures described in the *Publication Manual of the American Psychological Association*. Because of the great demand for journal space, authors should give special attention to reducing the length of manuscripts while maintaining clarity of expression. Ordinarily, papers should not exceed 15 typed pages. Accepted manuscripts are published in order of their receipt. An author may, however, request early publication of accepted manuscripts if he is willing to assume the cost of printing the necessary extra pages.

---

ARTHUR C. HOFFMAN
Production Manager

HELEN ORR
Promotion Manager

Communications including subscriptions, orders of back issues, and changes of address should be addressed to the American Psychological Association, 1333 Sixteenth Street, N.W., Washington 6, D. C. Address changes must reach the APA by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $10.00 (Foreign $10.50). Single copies, $2.00.

---

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Mt. Royal & Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-Class postage paid at Baltimore, Md.
Copyright, 1961 by the American Psychological Association, Inc.

JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 52 October 1961 Number 5

# PERSONALITY FACTORS AMONG SCIENCE AND TECHNOLOGY FRESHMEN[1]

EUGENE B. NADLER
*Case Institute of Technology*

AND

GILBERT K. KRULEE
*Northwestern University*

In spite of numerous criticisms of the concept of personality types, ably summarized by Hilgard (1957, pp. 485–488), the notion has proved useful in psychology and anthropology under the names of syndrome, life style, basic personality structure, modal personality, etc. (See Inkeles & Levinson, 1954, for example). The study reported here had as its starting point such a conception, one which could be used to differentiate among freshman students in a school of science and technology.

The focus was upon two dichotomous personality variables distinguished by Murray (1938, pp. 222–223), Extraception-Intraception and Exocathection-Endocathection. The first refers to the direction of the person's interest, outward or inward, while the second refers to its objects, things, or ideas. Taking all combinations of both the direction and the object of interest led Murray to describe the following four types:

1. Extraception-Exocathection: To adapt to the world as it stands; to be interested in tangible results; to be very practical; to amass a fortune. To secure a permanent position; to become a member of clubs and institutions; to be without illusions; to conserve established values. To work effectively with mechanical appliances.

2. Intraception-Exocathection: To live imaginatively; to dramatize the self; to express one's sentiments and beliefs in action. To initiate and further progressive social movements. To speak against abuses; to propose reforms. To concoct new schemes, business ventures, political innovations; to be guided by a vision of the future. To seek adventure; to become involved in amorous affairs.

3. Extraception-Endocathection: To be interested in ideas and theories about substantial events (for example, physical sciences). To reflect and write about external occurrences and systems; history, economics, government, education. To collect data and think inductively.

4. Intraception-Endocathection: To devote oneself to artistic or religious representations. To dream, brood, and introspect; to become absorbed in the attempt to resolve inner conflicts and spiritual dilemmas.

[1] This research was part of a larger series of studies on curriculum evaluation supported by a grant from the Carnegie Corporation of New York. Additional support was made available through a faculty research fellowship awarded to G. K. Krulee by the Social Science Research Council. The authors gratefully acknowledge the support and encouragement of the Case Institute of Technology administration. We are also very much in debt to Barbara Doty who assisted in the collection and analysis of the data and to Morrell Heald for his cooperation in the administration of the questionnaires. The opinions expressed herein are the authors'.

The study was concerned with the following tasks:

1. testing the hypothesis that significant numbers of students in the school would fall into the two extraceptive categories, and very few into the two intraceptive categories

2. placing the predicted finding in a broader societal context by determining what socioeconomic influences, if any, might be reflected in the two varieties of extraception

3. describing further the differences between the extraceptive students who were exocathecting on the one hand and endocathecting on the other in terms of their career goals and life values

4. examining the influence of these orientations on the students' evaluation of their courses and curricula, and on their academic performance

## Method

*Differentiating the types.* The following questions were included as part of a larger questionnaire[3] presented to 432 male freshmen about three-quarters of the way through their first semester:

1. Once in my life I would like to: a. develop a valid scientific theory, b. discover a satisfactory principle of ethical conduct, c. invent a marketable product, d. convince a legislature to enact one of my views into law.

2. I would like a job where: a. I could express my social views through my work, b. I could organize the thinking in my field in a systematic way, c. I could discover my true potentialities, d. I could do work of practical importance.

3. I am filled with awe when I consider the lives of: a. Vincent Van Gogh, Mahatma Gandhi, Albert Schweitzer; b. Charles Kettering, Andrew Carnegie, Winston Churchill; c. Louis Pasteur, Woodrow Wilson, Abraham Lincoln; d. Frank Lloyd Wright, John Dewey, Albert Einstein.

4. In my spare time I would like to: a. teach woodcraft to a Boy Scout troop, b. participate in a campaign for better schools, c. discuss and evaluate different political systems, d. learn about psychoanalysis.

5. Everyone is a little of everything, but by and large, I am mostly: a. practical, b. idealistic, c. thoughtful, d. sensitive.

For every question, each of the four alternatives corresponds to one of the four personality types. If in answering the five questions a student chose three or more alternatives characteristic of a single type, he was placed in that type. The chance probability of this happening was $p = .10$, calculated by collecting the first three terms of the binomial expansion $(\frac{1}{4} + \frac{3}{4})^5$.

*Questionnaire.* The same questionnaire included a large number of items which covered the general areas of social class background, career goals and life values, and evaluation of curricula and courses.

*Academic performance.* At the end of the first freshman semester, two measures of achievement became available, grade point averages and the number of separations.

## Results

Of the 432 students surveyed, 143 or 35% were classed as Extraceptive-Exocathecting, and 83 or 19% were classed as Extraceptive-Endocathecting. Only 15 men fell into the Intraceptive - Exocathecting category (3.4%), and only 9 into the Intraceptive-Endocathecting category (2.1%). The miscellaneous or mixed group (those with fewer than three responses in their largest category) consisted of 182 individuals, comprising 42% of the sample. Thus, the two Extraceptive groups who were predicted to make up a significant proportion of the sample together accounted for about 52% of the subjects, with only insignificant numbers in the Intraceptive categories. The rest of this report will be solely concerned with the two sub-

[3] A copy of this questionnaire has been deposited with the American Documentation Institute. Order Document No. 6834 from ADI Auxiliary Publication Service, Library of Congress; Washington 25, D. C., remitting in advance $2.00 for microfilm or $3.75 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

groups among the Extraceptive men, the Exocathecters and the Endocathecters, who will be referred to as the practical men and the theoretical men, respectively, for convenience and descriptive simplicity.

*Social Class Background*

The responses of the subjects to a question dealing with their fathers' occupations at the time of their birth (about 1940) and to one dealing with their fathers' occupations at the time of taking the questionnaire (early 1959) were sorted into the following categories: (*a*) unskilled, semiskilled, farm; (*b*) skilled; (*c*) foremen; (*d*) white collar, sales, government, military; (*e*) subprofessional; (*f*) entrepreneurial; (*g*) professional. Table 1 shows the percentage distributions of each group of fathers at the two times.

The chi square test of independence between the two distributions, practical versus theoretical fathers in 1940, was 15.97, significant at the .02 level. The differences between the two distributions were due to the higher proportion of unskilled, semiskilled, farm workers, and foremen among the fathers of the practical men, and the relatively higher proportion of white collar, sales, government, and military workers among the fathers of the theoretical men.

The distributions of fathers' occupations were still significantly different in 1959, chi square being 84.50, significant beyond the .001 level. However, the major difference here was due to the larger proportion of entrepreneurs among the fathers of the theoretical men.

Comparing each group with itself at the two different times showed a major shift among the practical fathers from unskilled, semiskilled, and farm in 1940 to white collar, sales, government, and military occupations in 1959; the chi square between the 1940 and 1959 distributions was 18.56, significant beyond the .01 level. The fathers of the theoretical men, on the other hand, showed a major increase in the entrepreneurial category; the chi square of 11.16 was not significant, but was close enough to significance to indicate a slight trend, $p$ being about .07. It is interesting that the occupational distribution of the fathers of the practical men in 1959 is almost identical with that of the fathers of the theoretical men in 1940.

There were no significant differences in fathers' incomes between the two groups, and the omission of almost 40% of the responses on questions dealing with parental education invalidated any conclusions that might have been drawn from these data.

TABLE 1

FATHER'S OCCUPATION AT DIFFERENT TIMES

| Father's occupation | Practical men (%) | | Theoretical men (%) | |
|---|---|---|---|---|
| | 1940 | 1959 | 1940 | 1959 |
| Unskilled, semiskilled, farm | 36 | 18 | 20 | 11 |
| Skilled | 11 | 13 | 16 | 19 |
| Foremen | 8 | 5 | 0 | 1 |
| White collar, sales, government, military | 18 | 35 | 29 | 30 |
| Subprofessional | 3 | 1 | 7 | 0 |
| Entrepreneurial | 8 | 12 | 12 | 26 |
| Professional | 16 | 18 | 17 | 14 |
| | $\chi^2$ | | $p$ | |
| Chi square comparisons: | | | | |
| Practical vs. theoretical: 1940 | 15.97 | | .02 | |
| Practical vs. theoretical: 1959 | 84.50 | | .001 | |
| Practical: 1940 vs. 1959 | 18.56 | | .01 | |
| Theoretical: 1940 vs. 1959 | 11.16 | | .07 | |

*Career Goals and Life Values*

In order to analyze the curriculum choices of the students, the school curricula were divided into science and engineering. The one curriculum which did not fit into either category, management, was safely excluded since only one practical man and one theoretical man chose it. The science curriculum comprised math, physics, chemistry, and engineering science. The engineering category included metallurgical, civil, chemical, electrical, and mechanical engineering. In this way a two by two table was constructed with personality type, practical or theoretical, on one margin, and curriculum choice, engineering or science, on the other. The chi square was 59.24, significant beyond the .001 level. Eighty-four percent of the practical men had chosen engineering, 16% science, as opposed to 67% of the theoretical men who had chosen science and 33% engineering.

TABLE 2

CHARACTERISTICS OF JOB AND LIFE GOALS CHOSEN AS IMPORTANT

| Characteristic of life goal | Practical men | Theoretical men |
|---|---|---|
| Ideal job: | | |
| Management | 17% | 4% |
| Equipment design | 14 | 4 |
| Graduate school | 30 | 73 |
| Characteristics of ideal job: | | |
| Use of special abilities | 62 | 81 |
| High income | 50 | 28 |
| Opportunity for creativity | 51 | 76 |
| Freedom from supervision | 30 | 44 |
| Ideal position at peak of career: | | |
| Science | 25 | 41 |
| Administration | 50 | 27 |
| Teaching | 1 | 9 |
| Personal characteristics thought important: | | |
| To be liked by people | 47 | 27 |
| To understand other people | 56 | 35 |
| Good grounding in basic theory | 59 | 73 |
| Practical knowledge of facts | 78 | 61 |
| Ability to convince and persuade | 40 | 24 |
| Organizing and administrative ability | 47 | 30 |
| Activities expected to give the most satisfaction: | | |
| Career | 22 | 36 |
| Family | 66 | 41 |
| Community and organizations | 2 | 10 |

The distribution of choices of the ideal job upon graduation for the practical men was significantly different from that of the theoretical men, chi square being 41.25, significant beyond the .001 level. The order of choices for the practical men was graduate school, management, equipment design, applied research, new product development, and bringing up the distant rear, basic research, production, sales, and teaching. The order of choices for the theoretical men was graduate school first, and all other alternatives far behind.

The requirements of the ideal job also differentiated the two groups. Out of a list of 10 possible requirements of the ideal job which were checked for "high" priority, four items discriminated between the two groups, the probabilities ranging between .05 and .001. Among the practical men it was more important to earn a good deal of money, whereas among the theoretical men it was more important to have the opportunity to use their special abilities and aptitudes, to be creative and original, and to be relatively free of supervision by others. Out of the same list which was rechecked for items having "medium" priority, five items separated the groups, the probabilities ranging between .05 and .02. Among the practical men it was moderately important to have the opportunity to use their special abilities and talents, to be creative and original, and to exercise leadership, whereas it

was moderately important for the theoretical men to have a chance to earn a good deal of money and to look forward to a stable, secure future.

Responses to a question about the occupational positions which would be ideal at the peak of their careers were classified into categories of scientific, administrative, ownership, teaching, and other and don't know. The chi square test of independence between the two distributions was 78.17, significant beyond the .001 level. Among the practical men the order of choices was administrative, scientific, ownership, and teaching. Among the theoretical men this order was scientific, administrative, teaching, and ownership.

The personal characteristics which the two groups perceived as promoting vocational success also showed some differences. Of 17 items checked as essential to vocational success, six distinguished the two groups at probability levels ranging from .05 to .01. The practical men felt it important to be liked by people, to understand other people, to be able to convince and persuade other people, to have practical knowledge of the facts in one's field, and to have organizing and administrative ability. The theoretical men more often felt it necessary to have a good grounding in basic theory.

A short series of four questions dealing with the importance of planning for the future, getting ahead in life, knowledge of one's academic performance in comparison to others, and choosing among being successful, independent, or well liked showed one significant difference between the two groups, significant at the .01 level. The practical men preferred success to independence, and independence to being well liked, whereas the theoretical men preferred independence to success, and success to being well liked.

The activities which were expected to give the most satisfaction in life showed the practical men overwhelmingly making family relationships a first choice, followed by career activities and leisure time activities. The theoretical men were about evenly divided between family relationships and career activities, with some sentiment for activities directed toward national and international betterment, and church activities. The chi square between the two distributions was 17.74, significant beyond the .01 level.

*Evaluation of Curricula and Courses*

In 5 out of 13 ratings describing the nature of the ideal curriculum, the two groups presented different conceptions; four of these differences were significant beyond the .001 level and one at the .01 level. The practical men thought the ideal curriculum should emphasize practical aspects of the material, relate directly to the kind of work they would be doing, emphasize facts and techniques that have direct application, leave enough free time for other activities, and prepare them for taking responsibilities in management and administration. The theoretical men thought that the ideal curriculum should be mostly concerned with theory, provide them with a broad and general knowledge of the field, emphasize the ability to think logically and abstractly, be difficult enough to require real effort, and not have to prepare them for taking responsibilities in management and administration.

Two out of 13 ratings describing what the students realistically expected of their curricula distinguished the two groups, one being significant at the .001 level and one at the .01 level. The theoretical men expected their chosen curriculum to be mostly concerned with theory, but did not

TABLE 3

MEAN RATINGS OF ITEMS RELATING TO THE CURRICULUM

| Item[a] | Practical men | Theoretical men |
|---|---|---|
| The ideal curriculum should: | | |
| 1. be mostly concerned with theory (1)[b] emphasize practical aspects (8) | 5.72 | 3.74 |
| 2. provide broad knowledge (1) relate directly to future work (8) | 4.15 | 3.15 |
| 5. emphasize the ability to think (1) emphasize facts and techniques (8) | 4.42 | 2.69 |
| 6. require real effort on my part (1) leave me with free time (8) | 3.78 | 2.57 |
| 10. prepare me for management (1) not prepare me for management (8) | 2.92 | 4.16 |
| My realistically expected curriculum would: | | |
| 1. be mostly concerned with theory (1) emphasize practical applications (8) | 4.80 | 3.32 |
| 10. prepare me for management (1) not prepare me for management (8) | 3.70 | 4.70 |
| In my least valuable course: | | |
| 1. practical applications are emphasized (1) the primary emphasis is on theory (8) | 5.10 | 3.62 |
| 3. students influence what is discussed (1) instructor decides what is discussed (8) | 5.57 | 6.86 |
| 11. the instructor is impartial (1) the instructor is not impartial (8) | 3.13 | 2.36 |

[a] All items are paraphrased.

[b] Numbers in parentheses are the extreme scale values.

expect it to prepare them for taking on responsibilities in management and administration; the opposite was true of the practical men.

Neither the nominations for the most valuable course, nor the descriptive ratings of the most valuable course distinguished the practical men from the theoretical men.

On the other hand, the nominations for least valuable course were different for the two groups, the chi square of 12.926 having a probability of less than .05. Whereas 44% of the theoretical men nominated graphics as the least valuable course, only 17% of the practical men did. On the other hand, 30% of the practical men chose chemistry as their least valuable course, as opposed to 24% of the theoretical men. Other courses were chosen with equal relative frequency by both groups. In 3 out of 14 ratings describing the least valuable course, the two groups presented different evaluations, each significant beyond the .005 level. The theoretical men described their least valuable course as primarily emphasizing practical aspects of the material and were more vehement in expressing the idea that the instructor did not allow the students to decide what to discuss in class. However, they also felt he was more impartial in dealing with individual students than did the practical men.

*Academic Performance*

The grade point average of the practical men at the end of the first semester was 1.034 on a 3-point grading system which included 2 negative quality points for F. The average of the theoretical men for the same period was 1.378. The difference between these two means was significant beyond the .01 level for a two-tailed test, $t$ being 2.991.

Since it was possible that the obtained difference in school achievement was due to differences in scholastic aptitude between the two groups, they were compared with respect to their performance on two tests of the

College Entrance Examination Board required of all freshmen. There was no significant difference between the two groups on the Math Aptitude test. On the Verbal Aptitude test, the theoretical men scored slightly higher than the practical men; $t$ was 1.83, significant at the .05 level for a one-tailed test but not for a two-tailed test. Thus, there was only weak support at best for the notion that the higher achievement of the theoretical men was due to higher college aptitude.

There was no significant difference in separations between the two groups at the end of the first semester, the percentage being 22% for the practical men and 20% for the theoretical men. However, on the hypothesis that personality compatibility or ego syntonicity (see Levinson, 1959) might play a role in the separation rate, the subjects were divided between those who had made compatible choices and those who had not. Practical men who chose engineering curricula, or theoretical men who chose science curricula were considered to have made compatible choices; practical men who chose science, and theoretical men who chose engineering were considered to have made incompatible choices. Of 174 men who had made compatible choices, 30 or 17% were separated at the end of the first freshman semester. Of 50 men who had made incompatible choices, 17 or 34% were separated at the end of the first freshman semester. The difference between these proportions was significant beyond the .01 level.

## Discussion

The major differences between these Extraceptive freshmen may perhaps be best understood in terms of the social psychological concepts of *status anxiety* and *status security*. The pattern of responses to the questionnaire, and especially the data on father's occupation, suggests that the practical men were relatively more anxious about their status, while the theoretical men were relatively more status secure. The orientations of the practical men tended to serve the ends of upward status mobility in modern industrial society; those of the theoretical men gave greater personal and horizontal mobility. The courses and curricula of the school were evaluated in terms of whether they met the requirements of these orientations.

In the days when the differences between science and engineering were sharp and clear these orientations might have had approximately equal viability. Boys who liked to work with their hands could go into engineering; thoughtful boys could go into science.

The trend in recent years, however, has been for the engineering curricula to become more like the science curricula, to emphasize basic scientific principles and broad theoretical skills. The first year curriculum facing the students in this study had this marked theoretical character, and was identical for all of them. As such, it was less ego syntonic for the practical men than for the theoretical men, which probably accounts for the lower grade point average of the practical men at the end of the first semester.

In the first year, all the practical men faced a program that was not ego syntonic. However, for those who had chosen an engineering curriculum, the first year was at least a stepping stone to a program that was more compatible with their needs. No such hope remained for those practical men or those theoretical men who had made incompatible curriculum choices, since the feedback from upperclassmen and faculty members regarding the nature of the second, third, and fourth year programs could only increase the conflicts arising from their incongruous decisions. Thus, at the end of the first

freshman semester, those who had made incompatible curriculum choices had double the separation rate of those who had not.

What practical recommendations may be made from these considerations?

In general, the specific recommendations tend to fall under two main headings: (*a*) techniques designed to help students make ego syntonic choices at the outset, and (*b*) techniques designed to make it easier to change from an incompatible program to one that is compatible.

The present state of personality measurement probably does not permit that the research tool used here to classify freshman subjects could also be used to select prospective freshmen, even if it were greatly improved. On the other hand, such an instrument would have great value in aiding the preliminary counseling of freshmen. Although the subjects of this study were not required to declare a choice of curriculum until the end of the first year, and had an opportunity to revise this choice even afterward, it is clear that sometime during the first semester most of the students had made a fairly binding decision, since almost all of them answered the question on curriculum choice, whereas other questions went unanswered. A guidance instrument would prove of value in clarifying for the student the consequences of given curriculum decisions, and could aid in reversing seemingly binding decisions which were incompatible with the student's interest.

As a further way of improving the initial choices of freshmen, the publicizing departments of technical schools could step up a campaign to present a more realistic picture of what a modern technical program entails, focusing on the emphasis on theoretical skills. The present study indicates that, by and large, there is a great deal of self-selection in the students who come to a technical school, and that, by and large, this self-selection is highly realistic. An informative publicizing campaign could improve an already effective self-selection system. It also might have the effect of attracting students who might otherwise invest their energies elsewhere, thus influencing another side of the selection process.

There remains the problem of changing incompatible choices once they have been made. Our impression is that the student's decisions have been accepted by members of his family and his friends to a high degree, so that part of his identity has come to rest on the expectations that he has induced in others. It would appear to be desirable for the school to counter this social support, to create a climate in which indecision, hesitancy, and ambivalence about curriculum is tolerated, in which shopping around is encouraged, and in which curriculum choices may be more easily postponed or changed.

A second recommendation which attacks this same problem is one which is already being carried out in some respects. With the breakdown of the traditional boundaries between science and engineering in the working world, there is some tendency for a greater number of course choices to be offered the student, thus increasing the possibility of accommodating both practical and theoretical interest patterns within the same department.

## Summary

Measuring Intraception-Extraception and Exocathection-Endocathection in the freshman class at a school of science and engineering showed almost no Intraceptive persons present.

The Exocathecting men tended to choose engineering as a major field of study, were oriented toward the application of social skills in managerial and administrative occupations as means of achieving money and success, and perceived family relationships as bringing satisfaction in life. The Endocathecting men tended to choose science as a major field of study, were oriented toward utilizing theoretical skills in graduate school and scientific occupations as means of achieving self-actualization, and perceived family and career activities as jointly bringing satisfaction in life. In general, the Exocathecting men seemed more anxious about their social status than the Endocathecting men. Compatibility between personal needs and objective features of the school situation was seen as a major factor in the higher grade point average of the Endocathecting men, and in the lower separation rate of those students who had made ego syntonic curriculum choices. Some practical recommendations stemming from these findings were briefly discussed.

## REFERENCES

HILGARD, E. R. *Introduction to psychology.* (2nd ed.) New York: Harcourt, Brace, 1957.

INKELES, A., & LEVINSON, D. J. National character: The study of modal personality and sociocultural systems. In G. Lindsey (Ed.), *Handbook of social psychology.* Cambridge, Mass.: Addison-Wesley, 1954.

LEVINSON, D. J. Role, personality, and social structure in the organizational setting. *J. abnorm. soc. Psychol.*, 1959, **58**, 171-180.

MURRAY, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1958.

(Received September 13, 1960)

*Journal of Educational Psychology*
1961, Vol. 52, No. 5, 232-235

# THE DRAW-A-MAN TEST IN INDONESIA[1]

R. MURRAY THOMAS AND ANWAR SJAH
*University of California, Santa Barbara* *Padjadjaran University, Bandung*

Indonesia, a nation of perhaps 90 million people, has no standardized intelligence scale. This state of affairs is the result of a variety of factors, including a dearth of trained test constructors, the existence of many different languages in the nation, and the impossibility of establishing accurate birthdates for many subjects who might form a standardization group.

In trying to provide a temporary instrument to use with children until an adequate test is developed, the writers tried the Goodenough Draw-A-Man scale on a sample of 1,245 children in Bandung, a modern city of about one million inhabitants in the mountains of West Java.

The results of this study are reported here as: (*a*) a comparison between the success of the Indonesian children and the success of the American children in Goodenough's original standardization group to discover whether American standards for the test should be transferred directly for use in Indonesia, (*b*) correlations between the Indonesian children's Draw-A-Man scores and their teachers' ratings of their general intelligence, and (*c*) correlations between the Indonesian pupils' Draw-A-Man scores and their marks in several school subjects to discover whether the test correlated with achievement in school subjects.

[1] This study was carried out under a teacher education project of the State University of New York financed by the Ford Foundation. The aid of Winarno Surachmad and research assistants Soeparman, Mariani Azis, Suharsini, and S. Hutabarat is gratefully acknowledged.

## METHOD

### *Subjects*

The 1,245 children were drawn from three sources: 639 children from five public elementary schools enrolling Grades 1–6, 524 childen from five private schools enrolling Grades 1–6, and 82 children from three private kindergartens (there are no public kindergartens).

The schools were selected by random sampling techniques from the Bandung district, and every child in class the days of the testing (late spring 1960) was included. In the final compilation of data 109 children were eliminated from the sample because their birthdates could not be established, and 28 were eliminated because they drew women or included cars or boats instead of drawing only a man. This reduced an original sample of 1,382 to a usable sample of 1,245.

The nation's ideal of compulsory, universal education has not been realized, primarily because of a shortage of school buildings. Thus not all Bandung children attend school. An inspection of the occupations of the fathers of school groups like this sample suggests that it is biased somewhat in favor of the middle and upper socioeconomic classes.

### *Procedure*

In each classroom the research assistant who administered all of the tests passed letterhead size paper to the subjects and asked them to draw, with pencils, a picture of a man. While the pupils drew, their teacher was asked to make a judgment of each

child's overall intelligence and to place him in one of seven graduated categories of intelligence. (The Indonesian term *ketjerdasan* was used to mean *intelligence*.) The investigators' purpose was to learn whether there was a relation between scores on the Draw-A-Man Test and intelligence, according to the teacher's own estimate of each pupil's overall ability. The seven intelligence categories were named: very highest, quite high, somewhat above average, average, somewhat below average, quite low, and very lowest. The teachers were instructed to think of these levels as referring to school childen; so the lowest category was for the least intelligent children found in school, not the least intelligent in the general population.

## Results

Table 1 compares the Indonesian pupils with the normal-school-progress group from Goodenough's original standardization sample. (Goodenough compared an advanced-progress group with normal-progress and retarded-progress groups but she based the test's standards only on the normal-school-progress group.)

Although the Indonesian data were at first separated for analysis into private and public school samples, the means and standard deviations for the private and public schools were found to be almost identical at each age level. Since it appeared that the two groups represented the same population, they were combined for comparison with the American sample.

As shown by Table 1, the Indonesian group's mean was above the Goodenough-normal-progress mean at ages 5 (difference significant at .01 level) and 6 (significant at .05 level). But beyond age 6 the Indonesian means lagged behind the Goodenough-normal-progress means at all ages from 7 through 12 (all significant at .01 level) with the exception of age 11 where the Indonesian group spurted ahead of the Goodenough group a slight but statistically insignificant amount. (At no age did the Indonesian average rise as high as the Goodenough - advanced - school - progress group or fall as low as the Goodenough - retarded - school - progress group.)

Indonesian scores on the Draw-A-Man Test at each age level showed significantly more variation than did the Goodenough - normal - progress scores. These differences in variances were significant at the .01 level for ages 6, 7, 8, 10, and 11. They were significant at the .05 level for ages 5, 9, and 12 (Table 1). Perhaps this is due to the fact that by definition the Goodenough-normal-progress group did not contain the extremes that would be found in an unstratified sampling of school children.

In Table 2 correlations ($r$) between Draw-A-Man scores and teachers' ratings of intelligence are reported for each grade level, kindergarten through sixth. The correlations were all positive and moderately low or, as in Grades 2 and 4, quite low.

Table 3 shows correlations ($r$) between pupils' Draw-A-Man Test scores and the most recent end-of-the-semester marks received by the pupils in various school subjects. (The Indonesian marking system uses numbers ranging from 1 to 10, with teachers also commonly adding marks halfway between each pair of numbers, such as 5½. The extreme marks of 1, 2, 9, and 10 almost never appear.)

All correlations were positive. They ranged from very low (.04) to rather high (.80). In general the highest correlations were between the Goode-

TABLE 1

COMPARISONS OF INDONESIAN WITH AMERICAN GROUP

| Group | Age at last birthday | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| Indonesian: | | | | | | | | | |
| *N* cases | 35 | 104 | 135 | 202 | 204 | 183 | 165 | 121 | 98 |
| *M* | 13.6 | 15.1 | 16.8 | 20.0 | 23.3 | 26.9 | 33.2 | 32.2 | 36.9 |
| Median | 12.4 | 14.7 | 16.3 | 19.5 | 22.4 | 24.9 | 34.8 | 32.7 | 37.5 |
| *SD* | 5.3 | 5.5 | 6.0 | 6.3 | 7.9 | 8.3 | 8.5 | 8.6 | 6.8 |
| Goodenough-normal-progress: | | | | | | | | | |
| *N* cases | 623 | 802 | 669 | 503 | 504 | 351 | 101 | 22 | |
| *M* | 10.2 | 13.9 | 18.5 | 22.5 | 26.2 | 30.6 | 33.0 | 35.2 | |
| Median | 9.9 | 13.9 | 17.8 | 22.3 | 25.8 | 30.7 | 33.1 | 36.1 | |
| *SD* | 4.1 | 4.4 | 4.7 | 5.4 | 7.0 | 6.6 | 6.6 | 6.2 | |

nough test and arithmetic marks. Next highest were those in physical science (not taught below Grade 5), geography, and biology. On the average, correlations with reading, Indonesian language (the national language), and Sundanese (the local dialect) were somewhat lower.

TABLE 2

CORRELATIONS OF GOODENOUGH TEST SCORES WITH TEACHER RATINGS OF INTELLIGENCE

| | Grade level | | | | | | |
|---|---|---|---|---|---|---|---|
| | *k* | 1 | 2 | 3 | 4 | 5 | 6 |
| *N* | 82 | 210 | 238 | 221 | 218 | 151 | 97 |
| *r* | .22* | .32** | .07 | .27** | .11 | .29** | .22* |

* $p < .05$.
** $p < .01$.

## SUMMARY AND CONCLUSIONS

In Bandung, Indonesia, the 1,245 school children, kindergarten through Grade 6, who were given the Goodenough Draw-A-Man Test did not, on the average, score as high as did the American group on which Goodenough originally based her standards for the test, suggesting that the American standards for mental ages should not be applied to Indonesians.

When Goodenough test scores were

TABLE 3

CORRELATIONS OF DRAW-A-MAN SCORES AND SUBJECT MATTER MARKS

| Subject matter | Grade level | | | | | |
|---|---|---|---|---|---|---|
| | 1 ($N$ = 210) | 2 ($N$ = 237) | 3 ($N$ = 217) | 4 ($N$ = 214) | 5 ($N$ = 158) | 6 ($N$ = 105) |
| Arithmetic | .42** | .52** | .42** | .55** | .56** | .80** |
| Reading | .43** | .22** | .54** | .29** | .22** | .20* |
| Indonesian | .07 | .31** | .22** | .15* | .24** | .32** |
| Sundanese | .26** | .06 | .25** | .35** | .27** | .04 |
| Geography | | | .25** | .30** | .44** | .64** |
| Biological science | | | | .18** | .27** | .55** |
| Physical science | | | | | .44** | .69** |

* $p < .05$.
** $p < .01$.

correlated with school marks in various subject matter areas, the median correlation was about .30. The highest relationships were found between the Draw-A-Man Test and arithmetic marks (from $r = .42$ to $r = .80$). Correlations between Draw-A-Man scores and teachers' judgments of children's intelligence were low, ranging from .07 to .32 for different grades.

The writers concluded that the Goodenough test yielded sufficiently promising results to warrant its use as a measure of ability for Indonesian children (especially for estimating arithmetic and science abilities in upper elementary grades) until a more adequate test has been developed.

## REFERENCE

GOODENOUGH, FLORENCE L. *Measurement of intelligence by drawings.* Yonkers, New York: World Book, 1926.

(Received December 21, 1960)

*Journal of Educational Psychology*
1961, Vol. 52, No. 5, 236-240

# AMERICAN GRADUATE STUDENTS:

## SOME CHARACTERISTICS OF ASPIRING TEACHERS AND RESEARCHERS

DAVID GOTTLIEB

*Department of Sociology and Anthropology, Michigan State University*

In recent years, increasing attention has been paid to the phenomena of adult socialization as observed within American colleges and graduate schools (Becker & Strauss, 1956; Gottlieb, 1961; Jacob, 1957; Merton, Reader, & Kendall, 1957). While the ratio of speculation to empirical findings in this field is high, the available material would indicate that changes do occur in the values, attitudes, and behavior patterns of students in training and that these changes may at least in part be attributed to the social system of our educational institutions. Several investigators (Becker & Geer, 1958; Spaeth, 1959) have attempted to note how changes in the occupational values expressed by graduate students may be tied in with a theory of "anticipatory socialization" (Merton & Kitt, 1955), a process by which the individual adopts the values of a group to which he aspires but does not yet belong.

### *Purpose*

The research being reported here deals with a comparison of two groups of American graduate students: (*a*) the student who comes into the graduate school system with a preference for a career in either teaching or academic research and reports that there has been *no change* in his preference as he has moved through professional training, (*b*) the student who states that since he has been in graduate school either teaching or research has become more attractive as a postgraduate career choice. The major purpose of this study was to investigate how these students, who find themselves at different stages of socialization and career commitment, vary in respect to certain values and attitudes.

### Method

*Subjects.* The National Opinion Research Center's sample of graduate students from which the data for this analysis has been taken is a two-stage stratified cluster sample of the enrollment of American graduate schools which offered the PhD in the "traditional arts and sciences" in the autumn of 1958. A total of 2,842 students from 25 graduate schools participated in this survey.[1]

*Procedure.* In order to obtain some measure of the source and direction of career change each student was asked the following:

> Please rank the following in terms of your personal preference as a future occupation.
>
> a. Teaching undergraduates ____
> b. Teaching graduates ____
> c. Doing research in your field ____
> d. Academic administration ____
>
> Have your opinions on these alternatives changed since you first decided to go to graduate school? Please write the letter or letters of alternatives which seem more desirable to you now.

The responses to this set of questions were used as the basis for a change index. Persons who indicated that any alternative was more desirable to them are considered "changers"; persons who indicate no change in preference are considered "nonchangers." It was found that 41% of the graduate students in this

[1] The research reported here is part of a larger investigation of American graduate students conducted by the National Opinion Research Center. I am indebted to James A. Davis, Study Director, and to Joe L. Spaeth, Assistant Study Director, for their continued cooperation.

TABLE 1

CAREER PREFERENCES AND CHANGE STATUS[a]

| Present preferred career | Changing to teaching (%) | Changing to research (%) | No change (%) | N | Total % |
|---|---|---|---|---|---|
| College teaching | 19 | 19 | 62 | 888 | 100 |
| Research field | 10 | 26 | 64 | 936 | 100 |
| Other[b] | 8 | 17 | 75 | 812 | 100 |

[a] At this point our discussion of changers is limited only to those who report a change toward research or teaching. A third group of ambiguous changers include those who show a preference for both teaching and research, either of these in conjunction with administration, or some other alternatives in addition to those mentioned above.

[b] The "other" group includes students who wish to enter a professional practice, administration, secondary school teaching, parish work, etc.

sample do report a change of some kind in their career preferences.

## RESULTS AND DISCUSSION

Table 1 shows which students in terms of present career preference have changed and the direction of that change.

From Table 1 it can be seen that about a third of the students who plan a career in either teaching or research indicate that there has been a career interest change of some kind. Among the "others" there is less overall change —amounting to one out of every four subjects in that category. In respect to direction of change Table 1 shows: (*a*) those who state a current carreer preference for teaching are evenly split—19% stating a stronger interest in teaching and 19% a stronger interest in research, (*b*) 10% of the researchers are now more favorably disposed towards teaching while 26% state they now find research more attractive as a postgraduate career, (*c*) in the "other" category change goes in the direction of research at a better than 2 to 1 rate. The data suggest that the graduate school experience has the effect of bringing students closer to a research orientation, and that this move toward research is felt by students in each of the occupational groups.

### *Differences in Values and Attitudes*

Included in the graduate student questionnaire is a set of "dilemmas," a group of imaginary situations in which the respondent is asked, after being confronted with two lines of action with alternate implications: "What would you do?" or "What would you prefer?"

One dilemma has to do with the distinction between "localism" vs. "cosmopolitanism" in academic careers and consists of the following question: "In the long run would you rather be known and respected (*a*) throughout the institutions where you work or (*b*) among specialists in your field in different institutions?" Subjects who checked *a* were classified as "locals" while those who checked *b* were considered "cosmopolitans."

Table 2 shows that in each case

TABLE 2

CAREER PREFERENCE, CHANGE, AND "LOCAL-COSMOPOLITAN"

| Present preferred career | Cosmopolitan (%) | | | |
|---|---|---|---|---|
| | No change in career preference | N | Change in career preference | N |
| Teaching | 48 | 552 | 53 | 362 |
| Research* | 77 | 586 | 67 | 587 |

*$\chi^2$—significant at <.001 level (research, changers, nonchangers).

TABLE 3

CAREER PREFERENCE, CHANGE, AND WORK ORIENTATION

| Present preferred career | Percentage who would devote all or the bulk of their uncommitted time to work | | | |
|---|---|---|---|---|
| | No change in career preference | *N* | Change in career preference | *N* |
| Teaching* | 69 | 552 | 52 | 371 |
| Research | 69 | 554 | 64 | 590 |

*$\chi^2$—significant at <.001 level (teachers—changers and nonchangers).

those with the research orientation show the greatest desire for the national recognition. The difference between the changers and the nonchangers is found when the actual percentage differences are compared. The difference is one of intensity and not direction. Assuming that one's cosmopolitan orientation is tied to his career preferences, the results in Table 2 are what might be expected. Researchers who have not changed are most cosmopolitan as this is the type of professional view that would go with an unchanging attachment to a research career. Teachers who have not changed are least cosmopolitan since their frame of reference and operation would most likely be the local academic institution.

Significant differences were also found when respondents in each of the four groups were compared in respect to the *postgraduate* use of free time. Table 3 shows that while almost two-thirds of the research changers are inclined to devote the major part of their time to their work, only about half of the changers toward teaching are so inclined. A tentative interpretation would be that changers to teaching consist of a number of students who see teaching as a way of life in which the individual is able to regulate his time since he is guided by the number of hours a specific class meets. It may well be that these are graduate students who hold the more traditional (and perhaps naive) view of the academic world—one in which the professor prepares his lectures, meets his classes, and then retires to the solitude of his study. The nonchanging teachers appear to be no different than the researchers in the use which they would make of their free or uncommitted time. Career preference and degree of career commitment are also found to be related to how respondents view the teaching and research activity.

Table 4 shows that researchers are least likely to agree with the statement, "teaching is intrinsically more satisfying than research," those who have changed to teaching, and those teachers who have not changed being most in agreement with the statement. Not unlike the "local-cosmopolitan" situation it would appear that the responses of the changers to teaching and research is partially determined by the fact that these are students in transit. Included are: (*a*) students who have moved from teaching to research and now question the validity of the statement; (*b*) students who entered the graduate system with a preference for teaching and who have withstood to some extent the research influences of the graduate school, and

TABLE 4

CAREER PREFERENCE, CHANGE, AND TEACHING VERSUS RESEARCH

| Present career preference | Percentage agreeing that teaching is intrinsically more satisfying than research | | | |
|---|---|---|---|---|
| | No change in career preference | *N* | Change in career preference | *N* |
| Teaching | 64 | 543 | 56 | 363 |
| Research* | 7 | 580 | 26 | 570 |

*$\chi^2$—significant at <.001 level (researchers—changers and nonchangers).

in a sense have rededicated themselves to teaching; (*c*) those students who have made the shift from research to teaching. On the other hand, the nonchangers present a different picture. Here are people who have retained that which they brought with them into the system in respect to teaching and research. The nonchanging researchers are strongest in their convictions that in fact research is more satisfying than teaching. The teachers who have not changed are most willing to accept the idea that teaching does hold an intrinsic satisfaction over research.

Responses to yet another "dilemma" situation provide further evidence to support the proposition that there are significant differences in attitudes and values when comparisons are made between teaching and research oriented students. In this case respondents were asked the following:

> You are a full professor in a small but good department. Two men are being considered for a tenure position, but only one can be appointed.
>
> *Smith* is very well liked by the students, faculty, and throughout the institution. He is an effective teacher and is extremely well read, both in his field and in other areas.
>
> It is quite clear, however, that he is never going to produce research or work of scholarship which will make a fundamental contribution in the field.
>
> *Jones* is known to everyone as a *prima donna*. He is difficult to get along with, and students complain about his lack of interest in teaching. It is quite clear, however, that in the long run he is destined to be one of the major intellectual figures in the field.
>
> Which man would you vote for when the department makes its recommendations?

Each group, it would appear, is most eager to give support to his own kind. The teachers, be they changers or nonchangers, are most enthusiastic about Smith. The researchers, especially those who report no change, are least interested in bringing Smith into the department. While Table 5 does support a general hypothesis of differences among teachers and researchers, it presents data of additional interest. While each group indicates a preference for a colleague who is most like themselves, it will be noted that in each cell the majority express a preference for *Smith*—the teacher. It is possible to only speculate why this is so. Perhaps individuals prefer Smith since he seems most congenial and easiest to get along with. On the other hand it may well be that researchers and teachers are expressing a need for something they find lacking in their own graduate experience, namely a faculty member who would give more of his time to his students and less to his research interests.

TABLE 5

CAREER PREFERENCE, CHANGE, AND TENURE DILEMMA

| Present career preference | Percentage who would choose Smith | | | |
|---|---|---|---|---|
| | No change in career preference | *N* | Change in career preference | *N* |
| Teaching | 88 | 552 | 85 | 370 |
| Research | 68 | 593 | 71 | 588 |

## SUMMARY

The results of a national survey of 2,842 graduate students would indicate that graduate students do alter their career preferences as they progress through their professional training. Significant differences are found in attitudes and values when these students are compared in respect to career preference and degree of career commitment.

The data presented would suggest that a form of "anticipatory socialization" is in operation and has the effect

of preparing students for their post-graduate careers.

## REFERENCES


BECKER, H. B., & GEER, D. The fate of idealism in medical school. *Amer. soc. Rev.* 1958, **23,** 50–56.

BECKER, H. B., & STRAUSS, A. Careers, personality and adult socialization. *Amer. J. Soc.*, 1956, **62,** 253–263.

GOTTLIEB, D. J., Processes of socialization in American graduate schools. *Soc. Forces*, 1961, **40,** in press.

JACOB, P. E. *Changing values in college*. New York: Harper, 1957.

MERTON, R. K., & KITT, A. S. Contributions to the theory of reference group behavior. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Continuities in social research: Studies in the scope and method of the American soldier*. Glencoe: Free Press, 1955.

MERTON, R. K., READER, G., & KENDALL, P. *The student physician: Introductory studies in the sociology of medical education*. Cambridge: Harvard Univer. Press, 1957.

SPAETH, J. L. Value orientations and career preferences of graduate students. Paper read at the Chicago meetings of the American Sociological Society, September 1959.


(Received February 8, 1961)

*Journal of Educational Psychology*
1961, Vol. 52, No. 5, 241-248

# THE ACQUISITION AND RETENTION OF PAIRED ASSOCIATES BY GOOD, AVERAGE, AND POOR READERS[1]

WAYNE OTTO
*University of Oregon*

The present study investigates the relative abilities of good, average, and poor readers to learn and to recall a list of paired associates. The particular variables selected for study in this connection were chosen primarily because of questions raised by an earlier, exploratory study.

The exploratory study employed a learning task similar to the one used in the present study: a list of five paired associates, each pair consisting of a common geometric form and a low association value trigram. The task discriminated between good and poor readers sufficiently well to suggest further experimentation in examining several specific questions raised by the study.

1. In the earlier study, the subjects' IQ scores varied considerably. With intelligence level differences minimized, would the demonstrated differences between good and poor readers' ability to master paired associates persist?

2. The exploratory study had employed good readers and extremely disabled readers. Would the differences persist if the poor reading group consisted of subjects reading in the lower quarter of their class? How would a group of "average" readers perform?

3. There appeared to be a tendency for visual-auditory reinforcement to result in more rapid learning than did purely auditory reinforcement. With more adequate controls, would a particular mode of reinforcement continue to be more effective than another? Or, with provision made for the study of a possible interaction between mode of reinforcement and reading level, might one mode of reinforcement prove to be better for poor readers while another would be better for good readers?

4. If grade placement of subjects were to be varied systematically would this grade placement variable interact significantly with reading level and/or mode of reinforcement? For example, might not the most effective mode of reinforcement vary from grade level to grade level?

5. What is the effect of these variables—reading level, grade placement-chronological age, and mode of reinforcement—upon total trials required to relearn the list after 24 hours? The results of the earlier study suggest that there would be no difference in number of relearning trials required by good and poor readers.

Studies have been reported that have been concerned with one or more of the above questions. Five general statements regarding these studies can be made here.

1. Level of intelligence appears to be a significant variable in associative learning. There appears to be an inverse relationship between IQ and trials required to master a list of paired associates (Gates, 1930).

2. Studies show that associative

[1] The study is based upon data collected for a doctoral thesis at the University of Wisconsin. The writer is grateful for helpful comments to his thesis committee: Theodore L. Harris, Chester W. Harris, and Virgil E. Herrick.

learning ability and reading ability are directly related, but there is little agreement as to why or to what extent this is so (Fildes, 1921; Gates, 1935; Hincks, 1926; Monroe, 1932; Raymond, 1955; Vernon, 1958).

3. In general, studies have failed to demonstrate a clear advantage for any one type of reinforcement. The usual types of reinforcement investigated have been auditory, visual, and tactile/kinesthetic, presented singly and in combination (Berman, 1939; Burt & Lewis, 1946; Kirk, 1933; Raymond, 1955; Rivkind, 1958; Stauffer, 1948; Williams, 1952).

4. Studies have not shown adequate control of the learning tasks employed to permit conclusive interpretations of results. The failure of many studies to demonstrate advantages of any particular mode of reinforcement may simply reflect shortcomings of the task employed rather than absence of effects of the task variables.

5. There has been no real concern with what might be important interactions. Studies have generally been concerned with but one level of reading ability, with a particular mode of reinforcement, with a general range of chronological ages or levels of ability at the possible cost of what might have been significant interactions between two or more of these variables.

The present study, then, attempts to control and study what appear to be pertinent variables and make allowance for consideration of possible significant interactions between selected variables. The study is an experimental investigation of relationships that, once demonstrated, can serve as bases for further research in learning to read situations.

Three hypotheses provide a focus for analysis of the data and subsequent discussion.

1. Good readers, average readers, and poor readers will differ in the number of trials required to learn a list of paired associates.

2. Mode of reinforcement will interact significantly with reading level and/or grade level.

3. Good, average, and poor readers will not differ in number of trials required to relearn a list of paired associates 24 hours after acquisition trials.

## Method

Three major variables were studied: reading level, grade placement-chronological age, and mode of reinforcement.

### *Subjects*

One hundred and eight subjects (four replications of the $3 \times 3 \times 3$ design), chosen from a total population of about 2,900 elementary school children from three schools, served in the experiment. All subjects met the following criteria:

*Intelligence.* A group test IQ score between 95 and 110. Scores were from the Otis Quick Scoring at Grade 2, and from forms of the California at Grades 4 and 6. The attempt was to hold intellectual capacity within the average range by eliminating high and low deviant scores. The "average" score was subject to corroboration by the classroom teacher.

*Reading level.* Reading test scores falling in one of three specified categories denoted good, average, or poor. Scores were from forms of the Metropolitan Achievement Tests. Raw test scores were converted into stanine equivalents, which divide the distribution into nine subdivisions of approximately the same range. Thus, operationally defined, good readers were those whose test scores fell in the seventh, eighth, or ninth stanines; average readers had scores falling in the fifth stanine; and poor readers had scores falling in the first, second, or third stanine. The fourth and sixth stanines were dropped out to avoid overlap between groups. The level to which a subject was assigned was subject to corroboration by the classroom teacher.

*Grade level.* Placement at second, fourth, or sixth grade. Subjects were chosen from three grade levels in order to study the possible effect of educational development-chronological age upon efficiency in the learning task. Chronological age range was limited to 12 months to eliminate any wide

deviations from the usual age at grade level.

A pool of childen who met the criteria was selected and four subjects were then assigned to each experimental condition. Since the possible effect of sex differences upon learning efficiency under any of the experimental conditions was not known, two boys and two girls were assigned to each experimental condition. Because no systematic differences were demonstrated on the basis of sex, this will not be further discussed.

## *Task*

*List.* The paired-associates list employed in the study evolved from the preliminary study done by the writer. Each pair consists of a common geometric form and a low association value CVC trigram from the Archer (1960) list.

The specific geometric forms were chosen because preliminary investigation showed that most elementary school children had a dominant name for each one (diamond, circle, triangle, star, square). Geometric form-trigram pairs appeared to be preferable to trigram-trigram pairs to avoid needlessly confusing the younger subjects and subjects with poor reading ability with two word-like stimuli in each pair.

The trigrams chosen had association values of 25% or less according to the Archer list. The five trigrams were chosen to employ the short sounds of the common vowels—*a, e, i, o, u*—without repeating any consonant in the list.

The actual list follows. (The common name of the form is given here in lieu of the form itself.) Diamond-fep, circle-miv, triangle-wuc, star-yad, square-gox. In those reinforcement conditions in which the subject received visual presentation of the trigrams, the trigrams were printed in lower case manuscript style.

Since it was felt that some basis for comparison of associations evoked by the individual forms and trigrams in the different groups of subjects would have value in interpreting the data, an association test was implemented in the study. This test and results will be alluded to, but space does not permit further elaboration here.

*Apparatus.* A Stone memory drum was employed in controlling presentation of forms. The original box encasing the revolving cylinder provided an opening .5″ × .5″ for presentation of stimuli. A slide opening was devised to permit the experimenter to expose 1″ × .5″ portions of the total opening. The stimulus sheet mounted on the cylinder was so arranged that five different orders of presentation were possible. The forms presented on the memory drum were ⅜″ high.

*Acquisition.* All testing was done in a room with only the experimenter and the subject present and with a minimum of distracting outside noise. The writer served as experimenter throughout the experiment.

Each form was presented for 4 seconds—at which point the subject was expected to say the form name—followed by a 4-second presentation of the form-trigram, followed by a 4-second rest. There was a 16-second intertrial rest. Pairs were first presented in serial order until the subject reached the criterion of one correct anticipation of form names for the entire list. Pairs were then presented in scrambled order to the same criterion. The total score was serial plus scrambled trials to criterion. This sequence of presentation was employed because immediate scrambled presentation proved to be disconcerting to the younger subjects. The scrambled presentation countered any sequential learning effect.

*Recall/relearning.* Subjects were asked to relearn the list after 24 hours. Pairs were presented in scrambled order in the same manner as in acquisition trials. The subjects continued to the criterion of one correct anticipation of form names for the entire list.

## *Reinforcement*

The term "mode of reinforcement" is used in the present paper to refer to the manner in which the correct form-trigram pairs were presented to the subjects. "Confirmation" might be an equally apt term. The actual procedure is described in the following paragraphs. The point to be made here is that the word "reinforcement" does not conform to the strictly conventional usage.

Three modes of reinforcement were employed. In practice the reinforcement was cumulative: auditory, visual-auditory, and kinesthetic-visual-auditory. For the sake of simplicity, however, the three modes of reinforcement will be denoted auditory, visual, and kinesthetic, respectively.

*Auditory.* The form was presented for 4 seconds, followed by a second 4-second presentation of the form accompanied by the experimenter's articulation of the appropriate trigram (form name).

*Visual.* The form was presented for 4 seconds, followed by a second 4-second pres-

TABLE 1
MEAN ACQUISITION TRIALS

| Grade | | | Reading level | | | Reinforcement | | |
|---|---|---|---|---|---|---|---|---|
| 2 | 4 | 6 | Good | Average | Poor | Auditory | Visual | Kinesthetic |
| 16.42 | 11.42 | 9.14 | 8.67 | 13.03 | 15.28 | 13.36 | 12.33 | 11.28 |

entation of the form accompanied by the experimenter's articulation of the appropriate trigram plus visual presentation of the trigram. For the visual presentation, the experimenter held a card bearing the printed trigram contiguous to, and to the right of, the form presented by the memory drum. Cards, 1" × 2", with ½" × ½" centered letters, approximately the size of the forms, printed in black were employed.

*Kinesthetic.* The form was presented for 4 seconds, followed by a second 4-second presentation of the form accompanied by the experimenter's articulation of the appropriate trigram plus a visual-kinesthetic presentation of the trigram. For the visual-kinesthetic presentation, the subject was presented a card bearing the form-trigram pair and asked to trace the trigram with his finger. Cards, 10" × 4", with 1¼" × 2" letters, a size which the subject could readily trace, printed in black were employed. Letters were lower case manuscript, of the type most familiar especially to the second grade subjects.

*Pretraining.* Each subject was presented a card bearing a sample pair of stimuli (heart-keb) before acquisition trials began. The subject gave his name for the heart-shaped figure and was told that he would be asked to learn a new name. The test forms were then presented and paired with the appropriate trigrams in the manner of the reinforcement to be given. Subjects to receive kinesthetic reinforcement practiced tracing the sample trigram quickly.

## RESULTS AND DISCUSSION

A comparison of mean IQ scores shows no significant differences between groups compared in subsequent analyses. It appears that the random assignment of the pooled subjects to experimental conditions was reasonably successful in eliminating between group IQ differences; therefore, between group differences do not appear to be attributable to IQ differences between groups.

*Acquisition.* Mean acquisition scores for the different groups for each of the three main variables are shown in Table 1. An analysis of variance (Edwards, 1950) of the total acquisition scores is presented in Table 2. There was little evidence of heterogeneity of variance (chi square = 30.72 for 26 $df$, $p > .20$). Each of the three main effects is significant at the .01 level. That differences in performance by grade level are dependent upon reading level of subjects is indicated by the significant grade × reading level interaction. That differences due to mode of reinforcement are dependent upon grade placement of subjects is indicated by the grade × reinforcement interaction. There is no significant reading level × reinforcement interaction, nor is there a significant triple interaction.

Table 3 presents results of Tukey gap tests (Tukey, 1949), which were

TABLE 2
ANALYSIS OF VARIANCE OF THE ACQUISITION TRIALS

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Grade (G) | 2 | 498.94 | 79.91** |
| Reading Level (L) | 2 | 406.735 | 65.14** |
| Reinforcement (R) | 2 | 39.065 | 6.25** |
| G × L | 4 | 34.347 | 5.50** |
| G × R | 4 | 30.975 | 4.96** |
| L × R | 4 | 5.16 | |
| G × L × R | 8 | 11.562 | 1.85 |
| Within | 81 | 6.244 | |

** $p < .01$.

TABLE 3
RESULTS OF TUKEY TEST OF SIGNIFICANT GAPS

| Grade | | | Reading level | | | Reinforcement | | |
|---|---|---|---|---|---|---|---|---|
| Grade | Mean trials | Gap | Level | Mean trials | Gap | Mode | Mean trials | Gap |
| 2 | 16.42 | | P | 15.28 | | A | 13.36 | |
| | | 5.00** | | | 2.25** | | | 1.03* |
| 4 | 11.42 | | A | 13.03 | | V | 12.33 | |
| | | 2.28** | | | 4.36** | | | 1.05* |
| 6 | 9.14 | | G | 8.67 | | K | 11.28*** | |

* $p < .10$.
** $p < .01$.
*** The auditory/kinesthetic between means gap is significant ($p < .01$).

run to determine where significant differences between group means existed within the three main effects. Significant differences in mean acquisition trials were demonstrated between each of the three grade levels and between each of the three reading levels. The differences between the mode of reinforcement mean acquisition trials are approaching the .05 level of significance for each mode of reinforcement, but only the difference between the auditory and visual mode of reinforcement meets the $<.05$ requirement. Further interpretation must, of course, consider the significant interactions that have been demonstrated.

Several of the questions and hypotheses can be examined in view of the acquisition data.

Hypothesis 1: Good, average, and poor readers will differ in the number of trials required to learn a list of paired associates.

TABLE 4
MEAN ACQUISITION TRIALS: GRADE × READING LEVEL

| Grade | Reading level | | |
|---|---|---|---|
| | Good | Average | Poor |
| 2 | 10.58 | 18.17 | 20.50 |
| 4 | 8.33 | 11.83 | 14.08 |
| 6 | 7.08 | 9.08 | 11.25 |

The hypothesis is accepted, with the restriction that the grade level of subjects tested will tend to influence the magnitude of differences demonstrated. Table 4 makes clear the extent of the influence of grade level upon differences between reading levels. Examination of the mean acquisition trials presented shows that poor readers in Grade 2 took almost twice as many trials to master the list as did good readers; whereas poor readers in Grade 6 took only 1.6 times as many trials as did good readers. Also, good readers at the three grade levels perform more nearly alike than do the poor readers. It is suggested that the task is not sufficiently difficult at the higher grade levels to preserve the marked dichotomy displayed at Grade 2.

Hypothesis 2: Mode of reinforcement will interact significantly with reading level and/or grade level.

The hypothesis is accepted: there is a significant grade × reinforcement interaction. Mode of reinforcement does not, however, interact with reading level, contrary to what seems to be a basic assumption when added cues are stressed in much remedial work.

The nature of the grade × reinforcement interaction is made clear by Table 5. It can be seen that the most

TABLE 5
MEAN ACQUISITION TRIALS: GRADE × REINFORCEMENT

| Grade | Reinforcement | | |
|---|---|---|---|
| | Auditory | Visual | Kinesthetic |
| 2 | 16.33 | 18.42 | 14.50 |
| 4 | 13.58 | 9.83 | 10.83 |
| 6 | 10.17 | 8.75 | 8.50 |

efficient mode of reinforcement (requiring the fewest trials) varies by grade level. With the second grade subjects, kinesthetic reinforcement appears to make for the fewest trials; whereas, visual reinforcement appears to be most efficient for the fourth grade subjects and kinesthetic and visual are almost equally effective for the sixth grade subjects. The implication seems to be that the function of different cues supplied by the several modes of reinforcement changes as subjects progress in grade level and chronological age. This may help to explain why gross studies often fail to show a particular mode of reinforcement to be better than others. In order to secure positive results in any predictable manner, one would need to control grade level-chronological age.

The answers to two of the questions raised by the earlier study are clear from the consideration of Hypothesis 1 and Hypothesis 2: (*a*) When intelligence is held within a stated range, will good, average, and poor readers differ in the number of trials required to learn a list of paired associates? (*b*) Will there be significant interactions between reading level, mode of reinforcement, and grade level? The answers to both questions are affirmative, subject to the restrictions pointed out in connection with the hypotheses.

The answers to two other questions are also affirmative: (*a*) Will the traditional modes of reinforcement significantly influence rate of learning? It has been shown that the particular mode of reinforcement that is most effective will vary from grade to grade. (*b*) Is grade level-chronological age a significant variable in determining the number of trials required to master a list of paired associates? The affirmative answer is restricted by the fact that differences in number of trials required at different grade levels will be influenced by the reading proficiency of the subjects and by the mode of reinforcement employed.

*Recall and relearning.* "Recall score" is the number of form names correctly anticipated 24 hours after the acquisition trials. "Relearning score" is the total number of trials required to enable the subject to correctly anticipate form names for the entire list presented in scrambled order.

Mean recall and relearning scores for the several groups by variable are presented in Table 6. It can be seen that in general the mean recall scores vary in the same manner as the mean relearning scores. For this reason and because relearning efficiency probably

TABLE 6
MEAN RECALL AND RELEARNING SCORES

| | Grade | | | Reading level | | | Reinforcement | | |
|---|---|---|---|---|---|---|---|---|---|
| | 2 | 4 | 6 | G | A | P | V | A | K |
| Recall | 2.86 | 3.17 | 3.39 | 3.56 | 2.72 | 3.14 | 3.17 | 3.0 | 3.25 |
| Relearning | 2.67 | 2.11 | 2.11 | 2.08 | 2.50 | 2.31 | 2.44 | 2.22 | 2.22 |

most adequately reflects retention, only the relearning scores were subjected to analysis of variance. This analysis is presented in Table 7. Only grade level was found to be significant. Table 6 shows that the mean relearning scores for Grades 4 and 6 are identical. Consequently, it can be concluded that the only significant differences in relearning scores are attributable to the fact that second grade subjects took more trials to relearn than did fourth and sixth grade subjects.

Hypothesis 3 and the related question can be examined in view of the relearning data.

Hypothesis 3: Good, average, and poor readers will not differ in the number of trials required to relearn a list of paired associates 24 hours after acquisition trials. And the question: What is the effect of the variables—reading level, grade placement, and mode of reinforcement—upon total trials required to relearn the list after 24 hours?

The hypothesis is accepted: there is no difference in relearning performance due to reading level of the subjects. The answer to the question is that grade placement of the subjects alone significantly effects relearning after 24 hours.

That there were no differences between good, average, and poor readers in trials required to relearn the list seems to be a significant finding. This suggests that given sufficient reinforcement the future performance of good, average, and poor readers would tend to be similar. Thus, while subjects might require different numbers of trials to master a task, they might be expected to attain about the same performance level. This is, of course, part of the rationale proposed by some for the use of "teaching machines" or "autoinstructional devices."

TABLE 7

ANALYSIS OF VARIANCE OF RELEARNING TRIALS

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Grade (G) | 2 | 3.70 | 5.69** |
| Level (L) | 2 | 1.57 | 2.42 |
| Reinforcement (R) | 2 | .97 | 1.49 |
| G × L | 4 | .46 | |
| G × R | 4 | .13 | |
| L × R | 4 | .24 | |
| G × L × R | 8 | .98 | 1.51 |
| Within | 81 | .65 | |

** $p < .01$.

*Association test.* The association test results indicated that the poor readers gave more associations to the individual stimuli of the paired-associates list than did the good readers. As pointed out, space limitations preclude discussion of the association test; but the implication appears to be that there is no clear-cut relationship between the number of associations evoked (fluency) and efficiency in the present learning task.

## SUMMARY AND CONCLUSIONS

A 3 × 3 × 3 factorial design was employed to study the effects of three levels of reading proficiency, three levels of grade placement (chronological age), and three modes of reinforcement upon acquisition of a list of paired associates. One hundred and eight childen whose IQ scores were in the average, 95–110, range provided data for the investigation.

The results warrant the following conclusions:

1. Good, average, and poor readers—in that order—require increasingly more trials to master a list of paired associates.

2. Sixth, fourth, and second graders—in that order—require an increasing number of trials to master a list of paired associates; but these differences

are significantly influenced by both the reading proficiency of the subjects involved and the mode of reinforcement employed.

3. Mode of reinforcement as a variable interacts significantly with grade level. Therefore, studies purporting to investigate the relative efficiency of different modes of reinforcement must recognize grade level of the subjects as an effective variable.

4. There is not a significant interaction between mode of reinforcement and reading level. This is contrary to the assumption in instructional practice that additional cues will be of value in a remedial teaching/learning situation.

5. Analysis of total relearning scores showed grade level to be the only significant variable. Thus, it appears that once they have mastered a list of paired associates, poor readers will retain the associations as well as good readers.

6. Analysis of the results of the association test indicates that the differences in ability to learn a list of paired associates demonstrated by good, average, and poor readers cannot be explained simply in terms of number of meanings or associations attached to the stimuli.

Two limitations seem obvious in considering the significance of the present study for the teaching/learning of reading: (*a*) The universality of the present results needs to be demonstrated. (*b*) The functional relationship of the present task to learning to read is not known. It is felt, however, that the present results can be of value in formulating hypotheses to be evaluated when reading tasks are employed in the experimental setup.

## REFERENCES


Archer, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

Berman, A. The influence of the kinesthetic factor in the perception of symbols in partial reading disability. *J. educ. Psychol.*, 1939, **30**, 187–198.

Burt, C., & Lewis, R. B., Teaching backward readers. *Brit. J. educ. Psychol.*, 1946, **16**, 116–132.

Edwards, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

Fildes, Lucy G. A psychological inquiry into the nature of the condition known as congenital word blindness. *Brain*, 1921, **44**, 286–307.

Gates, A. I. *Interest and ability in reading*. New York: Macmillan, 1930.

Gates, A. I. *The improvement of reading*. (Rev. ed.) New York: Macmillan, 1935.

Hincks, E. M. *Disability in reading and its relation to personality*. Cambridge: Harvard Univer. Press, 1926.

Kirk, S. A. The influence of manual tracing on the learning of simple words in the case of subnormal boys. *J. educ. Psychol.*, 1933, **24**, 525–535.

Monroe, Marion. *Children who cannot read*. Chicago: Univer. Chicago Press, 1932.

Raymond, Dorothy M. The performance of reading achievers on memory span and associative learning tests. *J. educ. Res.*, 1955, **48**, 455–465.

Rivkind, H. C. The development of a group technique in teaching word recognition to determine which of four methods is most effective with individual children. Unpublished doctoral dissertation, University of Florida, 1958.

Stauffer, R. G. Certain psychological manifestations of retarded readers. *J. educ. Res.*, 1948, **41**, 436–452.

Tukey, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, **5**, 99.

Vernon, M. D. *Backwardness in reading*. Cambridge: Cambridge Univer. Press, 1958.

Williams, Gertrude H. An investigation of the relative values of certain word-learning techniques for retarded readers. Unpublished doctoral dissertation, Temple University, 1952.


(Received March 8, 1961)

*Journal of Educational Psychology*
1961, Vol. 52, No. 5, 249-253

# THE CONTRIBUTION OF RESPONSE IN MENTAL MEASUREMENT

WILLAVENE WOLF AND JAMES B. STROUD
*Ohio State University* *State University of Iowa*

Around the turn of the century, psychologists were active in investigating the relationship between speed of response and criteria of intelligence, as seen, for example, in the work of Gilbert (1894) and Wissler (1901). Most of the early workers dealt with simple motor responses, as in reaction time studies, or at best with simple responses. There appears to be little support for the existence of a general speed of response factor, as Tinker (1931) has pointed out. Generally negligible relationships have been obtained between speed of response on simple tasks and measures of intelligence. When significant relationships have been obtained, they have generally been interpreted as a function of the character of the specific task employed.

Thus it appears that rate of response on a speed test—one on which rate of response is the main determinant of the score—is essentially unrelated to performance on measures of intelligence. It does not necessarily follow, however, that rate of response on a power test is unrelated to criteria of intelligence. A subject who successfully completes difficult intellectual tasks quickly may have more mental ability than one who completes them, but must do so slowly. In such situations, it is reasonable to suppose that power makes for speed, as Kuhlmann (1939) and Tinker (1931) argued.

In addition to a concern about this general problem, the writers were interested in two intelligence tests: Raven's Progressive Matrices (revised order, 1956) and the Leiter International Performance Scale—both nonverbal tests. Specifically, we were interested in investigating the relationships between rate of response to the individual items on these two tests and the total score (i.e., number right on the two tests, respectively), and certain independent measures. Some recent work with Progressive Matrices yielded rather low correlations between total scores and achievement measures, around .40 (Knief & Stroud, 1959). The form of the test and the method of administration make this an ideal test for use with the deaf and others handicapped in the use of a language. Obviously any improvement in its predictive power would be to the good. Some items from the Leiter were selected and used simply to further the interest in the general problem—the contribution of rate scores on a power test.

## Procedure

The subjects in this investigation were 102 fifth grade pupils in two small midwestern communities, representing the entire fifth grade enrollment in the two communities. To these pupils were administered the Progressive Matrices, one-half of the tests in the Leiter scale, the Iowa Tests of Basic Skills, and the Lorge-Thorndike Intelligence test, Verbal Battery, Level 3. The Leiter tests used, starting with Year VI, were as follows: VI, 2 and 3; VII, 2 and 4; VIII, 2 and 4; IX, 3 and 4; X, 2 and 3; XI, 1 and 4; XII, 2 and 4; XIII, 1 and 4; XIV, 1 and 3; XV, 1 and 4; XVI, 3 and 4; XVII, 2 and 3. Progressive Matrices and the Leiter scale were administered individually. The examiner timed the

response to each item by means of a stop watch. The only alteration made in the directions was the addition of the sentence, "Do the items as quickly as you can, trying to get all the answers right."

The mean Lorge-Thorndike IQ of the sample was 109. The *SD* was 14. The mean PM raw score was 27.85. The *SD* was 8.19. The distribution was approximately normal.

In addition to the number right score on Progressive Matrices and the mental age score on the Leiter scale, a time score, based upon the average time required to compile the items done correctly, was computed for each pupil. The MA value of each item used on the Leiter scale was doubled, inasmuch as only one-half of the items was used. Progressive Matrices is made up of items of ascending difficulty. Thus it seemed probable that the time scores on some parts of the scale might be more significant than those on other parts. The scale was divided into fourths, numerically, and average time scores were computed for each pupil on each fourth of the scale, making it possible to determine the relation between speed of response and the criterion measures for each fourth of the scale.

Time allowances on the Iowa Tests of Basic Skills are rather liberal, but the test does impose time limits. The number of subjects finishing each subtest was ascertained. When fewer than 90% of the subjects had finished a subtest, questions at the end were eliminated—back to a point reached by 90%.

## Results

As a first step in the analyses of the data, zero order correlations were computed (*a*) between average time score on the PM and Leiter tests and ITBS and Lorge-Thorndike scores and (*b*) between number right scores on the PM and Leiter (MAs) and ITBS and Lorge-Thorndike scores. The results are reported in Table 1. The correlations between Progressive Matrices and the Leiter scale on the one hand and ITBS and the Lorge-Thorndike test, on the other, are surprisingly low. Since in this investigation the two tests were not administered in the standard way, the coefficients reported in Table 1 should not be taken as representative of the predictive power of the tests.

In order to determine the effectiveness of combining the average time scores with the number right scores on the PM and Leiter tests in the prediction of the composite achievement scores and verbal intelligence test scores, multiple correlation procedures were utilized. Table 1 gives the multiple correlation coefficients for the variables involved. The composite ITBS scores and the Lorge-Thorndike test scores are regarded as the dependent variables, and average time scores and number right scores on the PM and Leiter as the independent variables. Beta weights were computed and tested for significance. They are also reported in Table 1. Combining, in multiple correlation, time and number right scores on the PM and Leiter tests increased these tests' prediction on achievement and verbal intelligence scores to about the same degree. All beta weights, except that involving the Leiter time and the Lorge-Thorndike scores, were significant at the .01 level. Thus, the addition of time scores to PM and Leiter number right scores may be considered to have increased significantly the correlations between these two tests and the ITBS and Lorge-Thorndike scores.

The multiple correlation analysis indicated that making time a part of the scores increased the predictive

value of both the PM and Leiter tests. While the Leiter test did not lend itself to further analysis, since the subtests attempted by the respective subjects varied, further analysis was undertaken on the Progressive Matrices. To determine whether or not time on any fourth of the PM was more important for predictive purposes than was time on any other fourth, the PM was divided into sets of 14 or 15 items and the average time scores on each set were correlated with the ITBS and Lorge-Thorndike scores. The fourths represent roughly four different levels of difficulty. Only 74 subjects earned scores on each of the four sets of items. Thus this analysis was based on 74 cases. Table 2 presents the zero order correlations obtained between the average time scores on each set of items and the ITBS and Lorge-Thorndike scores. The only significant correlation was that between average time for the third set of items on the PM and the Lorge-Thorndike test.

Multiple correlations were computed to determine whether or not time scores on any fourth of the items on the PM when combined with the total number right scores yielded higher *R*s than those of any other fourth. The results are reported in Table 3. Beta weights were computed and tested for significance. They are also reported in Table 3.

The *R*s involving the time scores of the third fourth of items are higher than those involving any other fourth. One might have predicted that the magnitude of the *R*s would have increased with the difficulty of the items; i.e., that those of each succeeding fourth of the items would have been greater than those of the preceding fourth. Indeed a tendency in this direction was evidenced for the first three fourths. It seemed likely that the items comprising the last fourth of the test would, for these fifth grade pupils, be quite difficult and that, therefore, the average time scores of the various pupils would be based upon relatively few correctly

TABLE 1

CORRELATIONS BETWEEN THE RAVEN AND LEITER TESTS AND ITBS AND LORGE-THORNDIKE TEST AND BETA WEIGHTS

| Variable | ITBS | β | L-T | β |
|---|---|---|---|---|
| Raven Time | −.174 | −.35** | −.178 | −.37** |
| Raven Right | .217* | .39** | .254** | .43** |
| Multiple | .380** | | .420** | |
| Leiter Time | .080 | −.36** | .078 | −.043 |
| Leiter MA | .291** | .56** | .357** | .680** |
| Multiple | .310** | | .460** | |

Note.—Raven Time vs. Raven Right, $r = -.465$; Leiter Time vs. Leiter Right, $r = -.750$.

* Significant at .05 level.

** Significant at .01 level.

TABLE 2

ZERO ORDER CORRELATIONS BETWEEN AVERAGE TIME SCORES FOR EACH FOURTH OF THE PROGRESSIVE MATRICES AND ITBS, LORGE-THORNDIKE, AND TOTAL NUMBER RIGHT ON PM

($N = 74$)

| Average time | ITBS | Lorge-Thorndike |
|---|---|---|
| All correct items on the first fourth of Progressive Matrices | −.175 | −.232 |
| All correct items on the second fourth of Progressive Matrices | −.178 | −.147 |
| All correct items on the third fourth of Progressive Matrices | −.231 | −.258* |
| All correct items on the last fourth of Progressive Matrices | −.104 | −.030 |

Note.—Raven Time on first 14 items vs. Raven Right, $r = .157$; Raven Time on second 15 items vs. Raven Right, $r = .330$; Raven Time on third 15 items vs. Raven Right, $r = -.458$**; Raven Time on fourth 15 items vs. Raven Right, $r = -.298$*.

* Significant at .05 level.

** Significant at .01 level.

TABLE 3

MULTIPLE CORRELATIONS INVOLVING EACH FOURTH OF PM ITEMS

| Variable | $R$ | $\beta$ | $\beta$ |
|---|---|---|---|
| 6.15 | .30* | 1 = −.21 | 5 = .19 |
| 6.25 | .37** | 2 = −.34** | 5 = .36** |
| 6.35 | .45** | 3 = −.46** | 5 = .45** |
| 6.45 | .33* | 4 = −.30 | 5 = .37** |
| 7.15 | .37** | 1 = −.28* | 5 = .30** |
| 7.25 | .38** | 2 = −.32* | 5 = .40** |
| 7.35 | .52** | 3 = −.52** | 5 = .52** |
| 7.45 | .32* | 4 = −.22 | 5 = .37** |

Note.—1. Average Time on first 14 items on Raven's PM.
2. Average Time on second 15 items on Raven's PM.
3. Average Time on third 15 items on Raven's PM.
4. Average Time on fourth 15 items on Raven's PM.
5. Number right on complete Raven's PM.
6. Iowa Tests of Basic Skills.
7. Lorge-Thorndike Intelligence Test.
* Significant at .05 level.
** Significant at .01 level.

done items. This, if true, should tend to lower the reliability of the average time scores for this fourth of the test. Accordingly, the number of items successfully completed in each fourth of the test, and upon which the average times were based, was ascertained. These, for the successive fourths, were found to be 12.7, 7.7, 6.5, and 2.1. Thus there is room for doubt that the apparent exception to the trend for the $r$'s to increase in magnitude seen at the last fourth of the test is genuine. When those pupils who got only one right or none, were eliminated, the mean number right was still only 2.8 on the last fourth of the test.

One additional treatment of the data for Progressive Matrices was undertaken. Median time scores were computed for each item and one point was given for each time score which was faster than the median score. These points for time were then added to the number right score on the PM for each pupil and a new score obtained. Scores thus derived were then correlated with the criterion measures. Coefficients of .35 and .39 were obtained between the new PM scores and ITBS and Lorge-Thorndike scores, respectively, as compared with .217 and .254 between the PM number right scores and the same two criterion measures.

The data suggests that speed of response when combined with number right on the PM and Leiter increases significantly the correlation between these tests and the ITBS and Lorge-Thorndike scores.

In one recent investigation an average $r$ of .41 was obtained between Progressive Matrices scores and various ITBS subtest scores, on 164 fourth grade pupils (Knief & Stroud, 1959). Most reported correlations that we have seen run somewhat higher than this. In the present investigation ($N$ = 102), an $r$ of .217 was obtained between PM scores (number right) and ITBS composite scores and an $r$ of .254 between PM scores and Lorge-Thorndike verbal scores. Leiter MAs (based upon one-half of the items) correlated with these same two criterion measures to the extent of .291 and .357. When multiple $R$s were computed between (*a*) PM number right and average time scores and (*b*) Leiter MA and average time scores and these same two criterion measures, values of .380 and .420 were obtained for PM and of .310 and .460 for the Leiter. A coefficient of −.465 was obtained between average time values and number right scores on PM, and of −.75 between Raven average time values and MAs.

What the results of this investigation tend to show is that the speed with which correct responses are made on the PM and Leiter tests, when combined with the conventional scores, makes a significant contribution to the predictive power of these two tests, as

signified by increased correlations with the criterion measures. It is true, of course, that the correlations are still quite low, but the investigation does point up the importance of speed of response in the measurement of intelligence. Thus the results support the proposition that a person who can perform mental feats, especially difficult ones, quickly should be judged to have more ability than one who does so slowly.


## Summary

To investigate speed of response as a factor in mental measurement, the Progressive Matrices and the Leiter International Performance Scale were administered in individual sessions to 102 fifth grade pupils, and a time score was determined for each pupil.

The utilization of the time values together with the conventional scores resulted in correlations between the two tests in question (PM and Leiter) and the two criterion tests, the Iowa Tests of Basic Skills and the Lorge-Thorndike, significantly higher than the correlations obtained between the conventional scores alone and the two criterion tests. Further analysis of the PM data suggested that the contribution of the time values increased from fourth to fourth of the items, and thus with difficulty, through the third fourth.

The results imply that the person who is able to deal more quickly with intelligence test items has more ability than the slower, but equally accurate individual.


## REFERENCES


Gilbert, J. A. Researches on the mental and physical development of school children. *Stud. Yale Psychol. Lab.*, 1894, **2**, 40–100.

Kniep, L. M., & Stroud, J. B. Intercorrelations among various intelligence, achievement, and social class scores. *J. educ. Psychol.*, 1959, **50**, 117–120.

Kuhlmann, F. *Tests of mental development.* Minneapolis: Educational Test Bureau, 1939.

Tinker, M. A. The significance of speed in test response. *Psychol. Rev.*, 1931, **38**, 450–454.

Wissler, C. The correlation of mental and physical tests. *Psychol. Monogr.*, 1901, **3**(6, Whole No. 16).


(Received March 15, 1961)

*Journal of Educational Psychology*
1961, Vol. 52, No. 5, 254-261

# EFFECTS OF TRAINING AND ANXIETY UPON TEACHERS' PREFERENCES FOR INFORMATION ABOUT STUDENTS[1]

PHILIP J. RUNKEL AND DORA E. DAMRIN
*Bureau of Educational Research, University of Illinois*

A characteristic of the individual's perceptual or cognitive space which captures much attention in current literature—e.g., Bieri (1955), Bieri and Blacker (1956), Leventhal (1957), Lundy and Berkowitz (1957)—is the "complexity-simplicity" of the frame of reference within which the individual perceives environmental objects. Complexity-simplicity can be conceived as the dimensionality of the cognitive space. By dimensionality we do not refer to the nature or content of the dimensions mediating the response, but only to the sheer *number* of the dimensions used by the individual in discriminating environmental objects. This paper examines two hypotheses: (*a*) that training is related to cognitive dimensionality, and (*b*) so also is anxiety. The tests of these hypotheses were set in the context of teachers' thinking about problems of students.

## *Dimensionality and Knowledge*

Let us consider a couple of common notions about training or knowledge. On the one hand, it is widely held that training serves to rule out irrelevancies. By focusing attention upon what is relevant and important, training *restricts* the number of criteria by which matters are judged. On the other hand, a proposition dear to educators is that training is *liberating*; it provides the individual with an enhanced repertoire of criteria from which he can select the most appropriate to guide his judgment. Whereas the untrained individual is more prone to adopt an irrelevant dimension or one given by his mood of the moment, the individual with sufficient training possesses a kit bag of relevant dimensions and can judiciously select from it dimensions suitable in different situations. From this point of view training *increases* the number of dimensions which, at any given moment, the individual can draw upon.

These two ideas are not contradictory. The first says that the untrained person's criteria or dimensions of judgment depend on his mood of the moment or upon surface irrelevancies in the situation, so that in meeting a *series of situations* he utilizes a large number of dimensions. Although he might react on the basis only of one or very few dimensions in any one situation, his dimensions over many situations would be numerous. The second phase begins when the individual acquires a little knowledge. He then looks for certain relevant and important characteristics no matter what the situation, so that the tally of dimensions he uses over a series of situations is small. Then with further training, the person comes to recognize more criteria as relevant to each situation, selecting certain ones as more important in some situations and others as more important in others; consequently the dimensionality of his judgments over situations increases.

These ideas would lead one to predict that the number of dimensions likely to be used *in a single situation*

[1] The research reported herein was performed pursuant to a contract with the Office of Education, United States Department of Health, Education, and Welfare.

would increase with training. Over a series of situations, however, dimensionality would first decrease and then increase as knowledge was acquired. The particular prediction made in terms of our data will be better understood if we first describe our instruments and method.

## Method

Teachers in 28 secondary schools in Illinois were given a questionnaire describing twelve imaginary problems which a teacher might encounter with students. For each problem, the teacher was asked to express his preferences among six types of information which he might seek in order to deal effectively with the problem. Problem 1, along with a condensation of the instructions given to respondents, looked like this:

*Problem 1*

Joan McCauley has not turned in a single assignment since the beginning of the year. You have warned her that failure to turn in her homework will result in her failing the course, but this has no effect. She sits and daydreams most of the time she is in class.

As a working guess, where would you say the source of this difficulty might lie? To check up on your working guess, what additional kinds of information would you seek? Please number, in order, the types of information (no more than three).

| Achievement in school subjects | Family and home life | Intelligence | Fears and worries | Interests and ambitions | Personality and character |
|---|---|---|---|---|---|
| ___ | ___ | ___ | ___ | ___ | ___ |

To illustrate further, three more of the problems follow:

Virginia Smith's parents want their daughter to go to college, and Virginia herself wants very much to go. However, she has no idea what to major in.

Jim has been sent to the principal's office by two different teachers for his rude and sarcastic behavior in their classes. His view of every instance is, "The teacher started it."

Glen's teachers report that the boy comes to school so tired every day that he is always falling asleep in their classes. The school physician reports that he is basically in good health.

Each teacher chose three of the types of information and ordered them 1-2-3 according to his preference, and he did this 12 times, once for each problem. Using the 12 rankings given by a teacher (one ranking of three things for each of the 12 problems), it was possible to attempt to "unfold" them into one dimension according to the technique of Coombs (1953).

Unidimensional unfolding is a method for ascertaining whether an obtained rank order (or a set of rank orders) might have been generated from an underlying order. For example, suppose that a man with a wife and two children finds that six houses are available on the market, one with a combination living and sleeping room, another with one bedroom, another with two bedrooms, and so on up to five bedrooms. He might most prefer the house with three bedrooms, next the one with four, and thirdly the one with five. When he becomes the father of a third child, he might then most prefer the house with four bedrooms, then the one with five, and thirdly the one with three bedrooms. Both these rankings of three of the houses out of six could have been generated from the underlying dimension of the number of bedrooms in the houses. But if a man most preferred the house with, say, two bedrooms, he could not (on the basis of the number of bedrooms alone) next prefer the house with five—his next choice would have to be either the house with three bedrooms or the house with one. Thus some responses in rank order form are "unfoldable" into the number of bedrooms dimension and some are not. A more extended discussion of unfolding can be found in Coombs (1953).

An interpretation of unfolding in terms of the materials used in the present study might be illustrated as follows. Suppose that a teacher judged student problems in terms of what we might loosely call a "hard-soft" dimension. For one problem the teacher might feel that very "hard" information was most suitable, choosing first intelligence, then achievement, then interests, then family life, etc. On another problem the teacher might feel that an intermediate type of information was most suitable, choosing first interests, then achievement, then family life, then intelligence, etc. Both these preference orders could unfold into the same dimension, as will be seen below.

The unfolding technique does not, of course, depend on guessing what basis the

TABLE 1

ILLUSTRATION OF HOW THE RESPONSES OF ONE SUBJECT TO THE 12 PROBLEMS FITTED INTO AN UNFOLDING SET

| Type of information | Possible preference orders (each column shows the response to some problem) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intelligence | 1 | 2 | 3 | | | | | | | | | |
| Interests and ambitions | 2 | 1 | 1 | 1 | 2 | 3 | | | | | | |
| Achievement in school subjects | 3 | 3 | 2 | 2 | 1 | 1 | 1 | 2 | 3 | 3 | | |
| Personality and character | | | 3 | 3 | 2 | 2 | 1 | 1 | 2 | 2 | 3 | 3 |
| Family and home life | | | | | | 3 | 3 | 2 | 1 | 1 | 1 | 2 |
| Fears and worries | | | | | | | | | 3 | 2 | 1 | |
| Number of problems to which subject gave response shown in column[a] | 1 | 2 | | 1 | 1 | | | 1 | 1 | | 2 | |

[a] Responses of the subject which would not unfold with the other nine in this set were: Interests-family-fears, family-interests-personality, and fears-intelligence-interests.

subject might be using in organizing his preferences. The technique is entirely objective. As an illustration, the responses of one of the subjects are shown in Table 1. The 1-2-3 entries in the body of the table show one set of preference-orders which could be generated by the "underlying" order of information types listed at the left. For any given underlying order of the information types, and under the instruction to order three out of the six, there are eight sets (the table shows only one) of 13 response patterns, each being a set which could be generated by the underlying order when the respondent chooses, in order, the types nearest his "ideal" type. Eight sets are possible because of the effects of relative distances among the types; this, as well as other details of the unfolding technique, is explained by Coombs (1953).

The response patterns given by our illustrative subject are indicated at the bottom of the table. The particular unfolding set represented in the table accounted for nine of the subject's responses, and this was the most that any unfolding set would accommodate.

If any one of a teacher's 12 rankings failed to unfold with his others, we knew that at least two dimensions were required to encompass the 12 responses. We then went on to assume that the fewer of the teacher's 12 responses that would unfold together, the higher his dimensionality was likely to be. This would not necessarily be true, of course. But we assumed that it was more likely to be true than not.

The maximum number of rankings unfoldable into a single dimension was taken directly as the individual's dimensionality "score." Since there were 12 problems, the dimensionality score could range from 1 to 12. A large number of unfoldable responses means *low* dimensionality, and vice versa. Actually, the obtained scores ranged from 3 (high dimensionality) to 10 (low dimensionality). If a teacher gave an unusable response to more than one problem, he was coded NA and not given a dimensionality score.

The unfolding technique thus gave us a rough index of whether the teacher was using relatively few or relatively more dimensions in thinking about the types of information he would need in dealing with the problems described to him.

As a measure of training influences, we used a 60-item test developed in the course of our larger study: a test of Knowledge and Interpretation of Tests (KIT). We did not want a measure of years of experience in dealing with information about students, since, as it is often phrased, 10 years' experience often consists of 1 year's experience repeated 10 times over. Nor did we want a direct measure of amount of exposure to training, since there are great individual differences in the degree to which training "takes," especially when assessed a year or more after the last exposure to training. Consequently, we wanted what might be called a measure of the residual effects of training.

Some of the types of information offered the respondent were types typically obtained through the use of tests. This circumstance, together with the fact that training in the use of information about students usually includes training in constructing and interpreting tests, persuaded us that for the purposes of our hypothesis a measure of knowledge about testing would suffice to index the extent to which the individual's training would affect his tendencies in the use of information about students. Consequently, scores on the KIT were distributed against the dimensionality scores to test the hypothesis that increased training is associated with different dimensionalities. Some representative items from the KIT are given below.

In counseling junior high school students a teacher should be *most* cautious in her interpretation of the information she obtains from:

A. vocational interest tests
B. study skills tests
C. reading comprehension tests
D. performance tests

A student scored at the 75th percentile on a standardized achievement test. This means that:
A. 75% of the norm group scored lower than he did
B. 75% of the norm group scored higher than he did
C. he answered 75% of the questions correctly
D. he is in the upper 25% of his own high school class

On a group intelligence test, which pupil would be most apt to receive an IQ that is much lower than his true ability?
A. Mary, whose parents speak only Spanish
B. John, who failed in all of his school subjects last year
C. Jane, who comes from a lower middle class home
D. Bill, who spends all of his time in athletics instead of on his school work

The question arises what *kind* of relation should be predicted between our two measures—dimensionality (number of mutually unfoldable rankings) and training (score on the KIT). Can we specify precisely where, within the ranges covered by our measures, dimensionality should increase or decrease? We mentioned earlier that the predictions would be different if we were looking within a single situation or if we were looking at a series of situations. Was answering the Problems Questionnaire a single situation for our subjects, or was it a series of 12 situations? We could hardly know in advance.

Another difficulty in choosing a prediction was the range of the KIT. There was no way of determining in advance whether it extended high enough to reach the hypothesized point where dimensionality would begin to rise. Because of these uncertainties, we could not pick out a precise form of relation to predict. If the questionnaire were a single situation and the range of the KIT were too low, we could expect no relation at all to show up. If the questionnaire were a series of situations and the range of the KIT were moderate, we could expect dimensionality to decrease with knowledge. But if the KIT were ample in range, we could expect first a decrease in dimensionality and then an increase. We hoped that our instruments were adequate, and adopted the latter relation as our Hypothesis 1: Dimensionality in responses to the Problems Questionnaire will first decrease and then increase with score on the KIT.

## Results

### *Hypothesis 1*

Dividing the KIT scores into six intervals containing as nearly equal numbers of respondents as possible and entering the dimensionality scores for each KIT interval, we obtained the results shown in Table 2; males and females are tabulated separately.

The means shown in Table 2 are plotted graphically in Figure 1. Both

TABLE 2

Frequencies of Dimensionality Scores at Successive KIT Levels

| Males only Dimensionality | KIT scores | | | | | | |
|---|---|---|---|---|---|---|---|
| | 14–23 | 24–26 | 27–29 | 30–32 | 33–36 | 37–55 | Total |
| High 4 | 2 | 4 | 3 | | 2 | 3 | 14 |
| 5 | 5 | 8 | 3 | 4 | 8 | 10 | 38 |
| 6 | 10 | 11 | 5 | 7 | 9 | 6 | 48 |
| 7 | 5 | 5 | 7 | 8 | 2 | 5 | 32 |
| 8 | | 3 | 4 | 4 | 3 | 2 | 16 |
| Low 9 | | 2 | 1 | 2 | | 4 | 9 |
| Totals | 22 | 33 | 23 | 25 | 24 | 30 | 157 |
| *M* | 5.8 | 6.0 | 6.4 | 6.7 | 5.8 | 5.8 | |

| Females only Dimensionality | KIT scores | | | | | | |
|---|---|---|---|---|---|---|---|
| | 10–22 | 23–26 | 27–29 | 30–31 | 32–34 | 35–46 | Total |
| High 3 | 1 | | | | | | 1 |
| 4 | 2 | 5 | | 2 | 3 | 1 | 13 |
| 5 | 9 | 5 | 3 | 6 | 5 | 6 | 34 |
| 6 | 3 | 10 | 6 | 7 | 11 | 12 | 49 |
| 7 | 2 | 3 | 8 | 4 | 2 | 4 | 23 |
| 8 | 1 | | 2 | 4 | 4 | 5 | 16 |
| 9 | | | 2 | | | | 2 |
| Low 10 | | | | 2 | 1 | | 3 |
| Totals | 18 | 23 | 21 | 25 | 26 | 28 | 141 |
| *M* | 5.3 | 5.5 | 6.7 | 6.4 | 6.1 | 6.3 | |

Note.—Chi square tests from tables collapsed to 3 × 3: Males $\chi^2 = 11.13$, $p < .026$; Females $\chi^2 = 15.80$, $p < .003$.

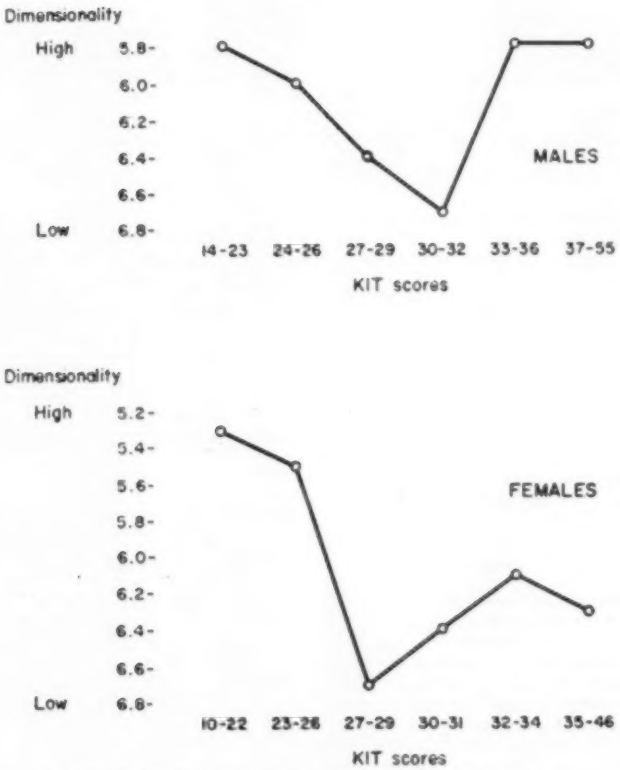


FIG. 1. Mean dimensionality at successive KIT levels.

curves dip in the middle with a gratifyingly high level of statistical confidence, and their general shape is replicated for males and females. Our theory suggests that the high dimensionality of the low KIT scorers represents a haphazard hopping from one dimension to another in going from problem to problem, whereas the moderate to high dimensionality of the high KIT scorers—so let us flatter ourselves—is the differentiated cognitive structure of the expert. The relatively low dimensionality of respondents with average knowledge might be called the simple sophomoric world of the semi-educated.

*Anxiety and Dimensionality*

There is evidence in the literature that anxiety focuses or narrows the aspects of the environment to which the organism responds. Easterbrook (1959) is particularly cogent on this point. Now, one way in which perception or attention can be focused is by reducing the number of dimensions along which environmental objects are discriminated; and our dimensionality measure is meant to indicate, roughly, just this: that the types of information presented to the teacher were discriminated along more or fewer dimensions. We should therefore predict (Hypothesis 2) that teachers with higher anx-

iety would show lower dimensionality than those with lower anxiety when responding to the problems.

Our measure of anxiety was the "Self-Perception Checklist" which formed part of a questionnaire used in our larger study. It was made up of selected items from the Guilford-Zimmerman Temperament Survey (1949)[2] and from Saunders' Personality Research Inventory (1955). After the returns were in, these items were analyzed for homogeneity. Thirteen proved to have tetrachoric $r$'s with total score of .40 or better, and these were used to make up the total score on anxiety. This score was merely the count of items answered according to the key for "security" as opposed to anxiety, high scores indicating security and the low scores anxiety, the two terms being meant to label a polarity. This short test was admittedly an unreliable instrument and subject to various faults such as social desirability answers; nevertheless, it was the best compromise we could make at the time with costs of administration, respondent fatigue, etc. Some examples of items are these:

Do you frequently get into a state of tension and turmoil by thinking over the day's happenings?

Have you usually been optimistic about your future?

Are you able to criticize people without hurting their feelings?

*Hypothesis 2*

When we distributed the security scores against dimensionality with all teachers pooled the result was about as near random as one could get. However, since neither of these measures was expressly designed to give a powerful test of the hypothesis we decided to look further for ways of selecting respondents for whom the security and dimensionality measures would be more discriminating.

Other work carried out by Gerald Larson (unpublished) as a part of the larger study had suggested that our security measure was more valid for males than for females. Also, as Table 2 shows, the KIT was not only associated with differences in dimensionality, but the spread of dimensionality scores was somewhat greater for the middle group of KIT scores. Accordingly, we distributed security against dimensionality for each of the six groups formed by sex and three levels of KIT. The results for males are shown in Table 3; the results for females were nowhere significant, and are not reported.

For males, results remained random at low and high levels of KIT but became significant in the middle range. In brief, no subgroup of respondents showed a relation contrary to hypothesis; and one subgroup, selected for sensitivity to our unreliable measures, showed the predicted relation at a good level of confidence. As the center subtable of Table 3 shows, relatively more of the secure male teachers exhibited high dimensionality and relatively more of the insecure showed low dimensionality.

*Security vs. KIT*

Because both knowledge and anxiety showed some relationship to dimensionality we investigated their relationship with each other. Only random results were obtained for females, but for males the relationship was monotonic. This result is reported in Table 4, where it is shown that more knowledgeable males tend to be more secure than do less knowledgeable ones. Whether the more secure males were those who took courses in testing and learned more from them or whether

[2] Used by permission. We used items from the Guilford-Zimmerman which had been selected in 1954 by Jay M. Jackson.

TABLE 3

DIMENSIONALITY VS. SECURITY WITHIN KIT LEVELS

Males Only

| Dimensionality | Security: Low (0-10) | Security: High (11-13) | Total |
|---|---|---|---|
| **KIT of 10–25** | | | |
| High 4–5 | 10 | 5 | 15 |
| 6 | 7 | 7 | 14 |
| Low 7–9 | 5 | 5 | 10 |
| Total | 22 | 17 | 39 |
| $x^2$ corrected for continuity = 0.32; *ns.* | | | |
| **KIT of 26–35** | | | |
| High 4–5 | 7 | *11* | 18 |
| 6 | 4 | *15* | 19 |
| Low 7–9 | *20* | 12 | 32 |
| Total | 31 | 38 | 69 |
| $x^2 = 8.54$, $p < .01$ (one-tailed). Italicized cell frequencies are higher than randomly expected. | | | |
| **KIT of 36–54** | | | |
| High 4–5 | 4 | 10 | 14 |
| 6 | 5 | 5 | 10 |
| Low 7–9 | 2 | 9 | 11 |
| Total | 11 | 24 | 35 |
| $x^2$ corrected for continuity = 1.33; *ns.* | | | |

they became more secure after acquiring more knowledge about testing is, of course, undetermined.

## SUMMARY

This paper is concerned with the effect of knowledge and of anxiety upon certain aspects of cognitive dimensionality. Dimensionality is here defined as the total number of dimensions or preference factors a person employs in perceiving and evaluating specified complexities of environmental stimuli. Individuals with high dimensionality employ a sizable number of dimensions in dealing with such stimuli; those with low dimensionality employ relatively few.

In this research the stimuli consisted of six general kinds of information a teacher might select for use in working with students who presented problems typical of those frequently encountered in the classroom. Our aim was to investigate whether differences in teachers' knowledge about such matters, and also whether differences in teachers' personal feelings of security, affected cognitive dimensionality in ways consistent with those theorized in current literature.

Hypotheses long held by educators combine to assert that training, by serving to increase knowledge, at first *restricts* and later *enlarges* the individual's cognitive framework. The untrained person, having no established framework, gives responses primarily determined by the situation. Training first serves to decrease this variability by providing a few specific ways of perceiving these stimuli regardless of the situation, and later serves to enrich cognitive dimensionality by providing the person with knowledge that equips him with more criteria for discriminating the stimuli. The signifi-

TABLE 4

SECURITY VS. KIT

Males Only

| KIT scores | Security: 0-7 | 8-9 | 10-11 | 12 | 13 | Total |
|---|---|---|---|---|---|---|
| 36–54 | 2 | 5 | 11 | 9 | 8 | 35 |
| 26–35 | 10 | 15 | 12 | 21 | 11 | 69 |
| 10–25 | 10 | 7 | 10 | 8 | 4 | 39 |
| Total | 22 | 27 | 33 | 38 | 23 | 143 |

Note.—Critical ratio by J. E. K. Smith's (1956) monotonicity test is 2.35; $p < .02$ (two-tailed).

cant curvilinear relationship found between teachers' knowledge of testing and teachers' perceptions of student problems strongly supported this hypothesis. Teachers with low and high levels of knowledge possessed high dimensionality: a finding we interpreted as indicating erratic versus specialized and differentiated perceptions. By contrast, teachers with average knowledge possessed relatively low dimensionality, which we interpreted as indicative of the restricted simplicity of the partially trained individual.

Another body of theory states that feelings of anxiety serve to narrow perception to a few elements of the environment. We reasoned that, if true, such narrowing would be reflected in a reduction of the number of dimensions the anxious individual employs in dealing with his environment. This hypothesis was supported by a part of our data. No relation between anxiety and dimensionality was found for female teachers, but a significant one was found for male teachers who possessed knowledge in the middle or average range. We believe that more reliable measures of anxiety and of dimensionality might provide stronger support for this second hypothesis.

## REFERENCES


BIERI, J. Cognitive complexity-simplicity and predictive behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 263–268.

BIERI, J., & BLACKER, E. The generality of cognitive complexity in the perception of people and inkblots. *J. abnorm. soc. Psychol.*, 1956, **53**, 112–117.

COOMBS, C. H. Theory and methods of social measurement. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden, 1953. Pp. 471–535.

EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychol. Rev.*, 1959, **66**, 183–201.

GUILFORD, J. P., & ZIMMERMAN, W. *Guilford-Zimmerman Temperament Survey*. Beverly Hills, Calif.: Sheridan Supply, 1949.

LEVENTHAL, H. Cognitive processes and interpersonal predictions. *J. abnorm. soc. Psychol.*, 1957, **55**, 176–180.

LUNDY, R. M., & BERKOWITZ, L. Cognitive complexity and assimilative projection in attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 34–37.

SAUNDERS, D. R. *The Personality Research Inventory*. Princeton: Educational Testing Service, 1955.

SMITH, J. E. K. *On the analysis of contingency tables with ordered classifications*. Lexington, Mass.: Massachusetts Institute of Technology, Lincoln Laboratory, 1956.


(Received March 29, 1961)

*Journal of Educational Psychology*
1961, Vol. 52, No. 5, 262-265

# LEARNING IN FOURTH AND SIXTH GRADERS AS A FUNCTION OF SENSORY MODE OF STIMULUS PRESENTATION AND OVERT OR COVERT PRACTICE[1]

JOAN LOCKARD AND JOSEPH B. SIDOWSKI
*University of Wisconsin* *San Diego State College*

A number of studies have indicated a relatively consistent trend for subjects to learn more efficiently when material is presented simultaneously to two sensory channels (auditory + visual) than when a single sensory input is used (Day & Beach, 1950). When auditory versus visual modalities are considered, however, results are conflicting. Interpretation of the results is further complicated by the interaction of sensory input variables with other variables. Early studies indicate that auditory presentation of material is more effective than visual for relatively young children (third, fourth, and fifth grades) but that the trend reverses as age increases (Erickson & King, 1917; Russell, 1938). Several recent studies (Evans, Glaser, & Homme, 1960; Silverman & Alter, 1960), investigating response modes in automated instruction with college students, have suggested that the requirement to write a response may be less effective in learning than merely reading the same material. The present study, therefore, was designed in order to investigate the influence of three modes of stimulus presentation (auditory, visual, and auditory + visual) and two modes of responding (overt or covert) upon learning in fourth and sixth grade children.

[1] The authors wish to express their sincere appreciation to Robert Nardelli, Principal of the Campus Laboratory School, San Diego State College, for his cooperation during the course of the experiment.

## METHOD

*Subjects.* Eighteen fourth and 18 sixth grade children from the Campus Laboratory School at San Diego State College served as subjects. The mean IQ (California Tests for Mental Maturity) for the fourth grade was 119.7, and for the sixth grade, 124.6.

*Procedure.* Students in both grades were required to learn lists of nonsense syllables under each of three sensory modes of stimulus presentation: visual, auditory, and visual + auditory. The subjects in each grade level were randomly assigned to two response subgroups, overt or covert, consisting of nine subjects each. In the overt response condition, subjects were required to recall and write down the syllables after each of 14 trials. The covert groups merely learned the syllables and were given a written recall test only after the fourteenth trial. A trial was defined as one complete presentation of the list of 10 syllables.

Five lists of nonsense syllables, two practice and three experimental, taken from Glaze (1928) and of high association value (87–100%) were used. In order to keep intralist similarity at a minimum, a different consonant started each of the syllables in a list and vowels were used about equally often. Each list consisted of 10 syllables. An earlier study by Sidowski and Eason (1960) indicated no significant difference in difficulty between the lists. Each list was presented for a total of 14 trials in each condition. Syllables were presented at a 2-second rate with a 2-second interval intervening between the presentation of each syllable. The syllables were spelled out during the 2 seconds for the auditory conditions by means of a Webcor tape recorder. Visual stimuli were presented at eye level by an American Optical Performance 500 projector at a distance of approximately 8 feet from the seated subject. The intertrial interval was 30 seconds for all conditions. Each subject in the overt groups was required to write down all of the nonsense syllables that he remembered during this interval on each of the 14 trials. Subjects in the covert groups merely sat and waited for the

next trial. All subjects were instructed to remember as many of the syllables as possible since they would be tested after going through the list 14 times. After Trial 14, subjects in all groups were allowed as much time as was needed to recall and write down the syllables.

*Design.* Each subject served in the experiment one hour per day for 4 consecutive days. The first day was a practice day during which the subject learned one practice list auditorily and a second visually. On each of the following 3 days, learning was carried out under one of the three experimental sensory conditions. The presentation of the sensory variables and the three experimental lists were counterbalanced within the design, as were the presentations of the practice conditions.

At the conclusion of the experiment, the subject was required to rank the three sensory variables from the least preferred to the most preferred.

## Results and Discussion

The mean number of errors in recall are presented in Table 1 for all conditions. (An error was defined as an incorrect recall as well as a failure to recall.) An inspection of the data indicates a consistent superiority for the visual, and visual + auditory presentations for the overt practice condition. This superiority is consistent for both grade levels. The covert practice condition appears superior for the sixth graders with little differentiation at this grade level between the sensory variables.

An analysis of variance was carried out on the data for Trial 14 since recall measures were taken during this trial for the overt and covert conditions. The results of the analysis are presented in Table 2. A significant overall difference was found between the sensory modes of stimulus presentation, while the Grades × Response Type interaction showed an $F$ of 4.11, with an $F$ of 4.15 being required at the 5% level. Analysis of the sensory mode variable indicated that the auditory presentation of stimuli was significantly inferior in recall to visual but not significantly different from auditory + visual.

TABLE 1

Mean Number of Errors in Recall for Overt and Covert Response Conditions

| Stimuli | Response groups | | |
|---|---|---|---|
| | Overt | | Covert |
| | Trial 1 | Trial 14 | Trial 14 |
| Grade 4: | | | |
| Auditory | 8.2 | 2.8 | 3.3 |
| Visual | 7.3 | 0.8 | 2.1 |
| Auditory + Visual | 7.3 | 2.0 | 2.3 |
| Grade 6: | | | |
| Auditory | 8.3 | 3.2 | 0.07 |
| Visual | 6.6 | 1.2 | 0.07 |
| Auditory + Visual | 7.1 | 1.5 | 0.07 |

It was decided beforehand to carry out two additional statistical analyses. A separate analysis was carried out on the covert response groups and the results indicated: (*a*) fourth graders producing significantly more errors than sixth graders, (*b*) no significant difference between sensory modes for the sixth grade, and (*c*) for the fourth grade, auditory presentation producing significantly more errors ($p < .05$) than visual. A final analysis of variance was carried out on Trial 1 for the overt response condition. The results indicated a significant sensory mode effect ($F = 7.95$, $df = 2/32$, $p < .01$) with auditory presentation showing significantly more errors than either of the other sensory modes for both grade levels.

It is obvious that the results of the present study do not agree with those of earlier studies (Erickson & King, 1917; Russell, 1938) which indicate a superiority for the auditory method for younger children. In the earlier studies, however: (*a*) more meaningful material was used, (*b*) the audi-

TABLE 2

ANALYSIS OF VARIANCE FOR ERRORS IN RECALL ON TRIAL 14

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Grades (G) | 1 | 22.23 | 2.83 |
| Response types (RT) | 1 | 2.09 | |
| G × RT | 1 | 32.23 | 4.11 |
| Subjects within groups | 32 | 7.85 | |
| Sensory modes (SM) | 2 | 15.87 | 9.12* |
| SM × G | 2 | 0.84 | |
| SM × RT | 2 | 5.08 | 2.92 |
| SM × RT × G | 2 | 2.07 | 1.19 |
| Error | 64 | 1.74 | |
| Total | 107 | | |

* $p = .01$.

tory presentation of material consisted of class lecture and/or reading by a status figure, and (*c*) less stringent controls were exercised over the experimental environment. One of the advantages inherent in the use of the visual method for presenting nonsense syllables is that the letters for any three letter syllable are presented simultaneously. With the auditory method, each of the letters is spelled serially thus requiring the subject to organize the letters in the syllable while learning (Krawiec, 1946). Evidently the covert condition allowed the sixth graders to learn and organize the material better than the overt method for all modes of stimulus presentation. This was not true for the fourth graders, although the younger children had little difficulty in recalling the material as well as the sixth graders when written responses were required after each trial.

The results of the preference rankings of the three sensory modes paralleled those of the error score analyses. The visual and auditory + visual modes were preferred significantly more than the auditory mode, with little preferential difference being indicated between the two visual conditions.

## SUMMARY

The learning of nonsense syllables by fourth and sixth graders was investigated as a function of: (*a*) three modes of stimulus presentation (auditory, visual, and auditory + visual); and (*b*) two response types, overt (written) and covert (nonwritten).

All subjects were required to learn a list of nonsense syllables under each sensory mode of stimulus presentation. Each list was presented for 14 trials under all conditions. Half of the subjects in each grade recalled and wrote out as many of the syllables as possible after each of the 14 trials (overt group). The remaining subjects merely observed and learned the syllables for 14 trials with no written recall responses (covert). After the 14 trials, subjects in all groups were given as much time as was needed to recall and write out the syllables.

Analyses of the data indicated a significant superiority for the visual and auditory + visual methods of stimulus presentation. No significant difference was found between the fourth and sixth graders for the overt response condition. For the covert groups, however, the fourth graders recalled significantly fewer syllables than the sixth graders.

## REFERENCES

DAY, W. F., & BEACH, B. R. A survey of the research literature comparing the visual and auditory presentation of information. *USAF tech. Rep.*, 1950, No. 5921. Wright-Patterson Air Force Base, (Dayton, Ohio).

ELLIOT, F. R. Memory for visual, auditory and visual-auditory material. *Arch. Psychol., NY*, 1936, **29**, No. 199.

ERICKSON, C. I., & KING, I. A comparison of visual and oral presentation of lessons in the case of pupils from the third to the ninth grade. *Sch. Soc.*, 1917, **6**, 146–148.

EVANS, J. L., GLASER, R., & HOMME, L. E. A preliminary investigation of variation in the properties of verbal learning sequences of the "teaching machine" type. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning: A source book*. Washington, D. C.: National Education Association, 1960. Pp. 446–451.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

KRAWIEC, T. S. A comparison of learning and retention of materials presented visually and auditorily. *J. gen. Psychol.*, 1946, **34**, 179–195.

LUMLEY, F. H. Research in radio education at Ohio State University. In, *Education on the air: Fourth yearbook of the institute for education by radio*. Columbus: Ohio State University, 1933.

RUSSELL, R. D. A comparison of two methods of learning. *J. educ. Res.*, 1938, **18**, 235–239.

SIDOWSKI, J. B., & EASON, R. G. Drive, verbal performance, and muscle action potential. *J. exp. Psychol.*, 1960, **60**, 365–370.

SILVERMAN, R. E., & ALTER, M. Note on the response in teaching machine programs. *Psychol. Rep.*, 1960, **7**, 496.

(Received April 5, 1960)

*Journal of Educational Psychology*
1961, Vol. 52, No. 5, 266-274

# THE ROLE OF DISCRIMINABILITY IN MEANINGFUL VERBAL LEARNING AND RETENTION

DAVID P. AUSUBEL AND DONALD FITZGERALD
*Bureau of Educational Research, University of Illinois*

In a recent study (Ausubel, 1960), it was shown that introductory material at a high level of abstraction, generality, and inclusiveness (advance organizers) facilitates meaningful verbal learning and retention. By deliberately introducing relevant and appropriately inclusive subsuming concepts into cognitive structure, one provides helpful ideational scaffolding which enhances the incorporability and longevity of the more detailed material in the learning passage.

Advance organizers, however, ordinarily have two distinct functions that correspond in turn to two different aspects of the unfamiliarity of meaningful learning material. Sometimes, as in the above mentioned experiment, the new material is almost completely unfamiliar in the sense that cognitive structure is barren of even generally related concepts. Under these circumstances the purpose of the organizer is simply to provide ideational anchorage or scaffolding. More typically, however, the new learning material (e.g., Buddhist doctrines) is a variant of related, previously learned concepts (Christian doctrines) already established in cognitive structure. Here the role of the organizer is not only to provide optimal anchorage at an optimal level of inclusiveness, but also to increase the discriminability of the learning passage from analogous and often conflicting ideas in the learner's cognitive structure.

This second role of organizers is predicated on the assumption that if the distinguishing features of the new learning passage are not originally salient or readily discriminable from established ideas in cognitive structure, they can be adequately represented by the latter for memorial purposes, and hence would not persist as separately identifiable memories in their own right. It is assumed, in other words, that only discriminable categorical variants of previously learned concepts have long-term retention potentialities. Thus, if a *comparative* type of organizer could first delineate clearly, precisely, and explicitly the principal similarities and differences between the new learning passage (Buddhism) and existing, related concepts in cognitive structure (Christianity), it seems reasonable to suppose that the more detailed Buddhist ideas would be grasped later with fewer ambiguities, fewer competing meanings, and fewer misconceptions suggested by the learner's knowledge of Christianity; and that as these clearer, less confused Buddhist meanings interact with analogous Christianity meanings during the retention interval, they would be more likely to retain their identity. In this experiment, therefore, the value of a comparative organizer was tested by contrasting its effects on the retention of a Buddhism learning passage with both those of a nonideational (*historical*) introduction and those of a simple *expository* organizer.

It is hypothesized, first, that to the extent that the organizer is rendered discriminable from related concepts (Christianity) established in cognitive structure, and hence to the extent that it increases the discriminability of the

Buddhism learning passage from these Christianity concepts, it facilitates the learning and retention of the new Buddhist ideas. In addition, it is hypothesized, for analogous reasons, that the discriminability (and hence the learning and retention) of the Buddhism passage varies as a function of the clarity and stability of the learner's existing knowledge of Christianity, and that subjects with relatively unclear and unstable concepts of Christianity derive relatively more benefit from the organizers than do subjects with clear and stable concepts in this area of knowledge.

These hypotheses were suggested by the findings of a previous experiment (Ausubel & Blake, 1958) in which the learning and retention of a comparative Buddhism passage were contrasted with that of an expository Buddhism passage. However, for the following reasons, the organizer approach was adopted in this experiment, in preference to manipulating the learning material itself: First, organizers provide advance ideational scaffolding. Second, they provide the learner with a generalized overview of *all* of the major similarities and differences between the two bodies of ideas *before* he encounters the new concepts individually in more detailed and particularized form. Finally, they create an advance set in the learner to perceive similarities and differences, and, by avoiding overly explicit specification, encourage him *actively* to make his own differentiations in terms of his own particular sources of confusion.

The hypotheses advanced in this study apply only to meaningful learning material which, although unfamiliar to subjects, is relatable to long-established and relatively stable concepts in cognitive structure. They do not apply to rote learning, to completely unfamiliar learning material, or to learning tasks that can only be related to unstable or recently learned concepts. The organizers, furthermore, must consist of ideational material (both similarities and differences) at a high level of abstraction, generality, and inclusiveness, rather than constitute a simple summary or a mere listing of specific comparative points.

## Method

### *Subjects*

The experimental population consisted of predominantly senior undergraduate students (94 women and 61 men) in six sections of an educational psychology course at the University of Illinois. All subjects were enrolled in one of 10 teacher education curricula at the secondary school level. The experiment was conducted separately in each section as a required laboratory exercise, and was performed during regularly scheduled class hours. In order to maximize ego involvement, subjects were informed that after the data were processed their individual scores, as well as the class results, would be reported to them.

### *Learning Passages, Organizers, and Measuring Instruments*

The learning material used in this study was a specially prepared 2,500-word passage dealing with Buddhist concepts of God, immortality, soul, faith, salvation, morality, and responsibility. These concepts were elaborated in considerable detail. The passage only presented the significant *ideas* of Buddhism and ignored material on the life of Buddha, the history and geographical distribution of Buddhism, schools of Buddhism, and Buddhist ritual. Flesch analysis of the passage yielded an abstractness score of 19.19 (highly abstract) and a readability score of 42.46 (difficult).

The topic of Buddhism was chosen both because it was explicitly unfamiliar to undergraduate students, and because it dealt with variants of previously learned concepts (i.e., Christian doctrines) generally familiar to all of our subjects and, presumably, reasonably well-established in the cognitive structure of most of them. Both criteria were important since the main purpose of the study was to

ascertain whether advance organizers could facilitate the learning and retention of *unfamiliar* meaningful material by increasing the discriminability between the new material and *related* concepts already established in cognitive structure. The use of unfamiliar learning material also made it possible for all subjects to start from approximately the same base line in learning the passage. Empirical confirmation of the unfamiliarity of the Buddhism material was obtained when a group of comparable subjects, who had not studied the learning passage, made scores which, on the average, were only slightly and not significantly better than chance.

Still another advantage of the Buddhism material inhered in the fact that its learnability was relatively uninfluenced in our population by such factors as sex, field of specialization, and variability in relevant incidental experience. It is true that women subjects made significantly higher scores than men subjects on both the Buddhism and Christianity tests. This superiority, however, was not a function of differential background, interest, or motivation related to sex per se, but reflected significantly superior academic aptitude as measured by verbal score on the School and College Ability Test (SCAT).

Knowledge of the Buddhism material was tested 3 and 10 days after learning, by equivalent forms of a 45-item multiple-choice test with corrected split-half reliabilities of .80 and .79, respectively. The correlation of the 3- and 10-day retention scores was .79. This latter figure indicated both relatively high stability over time, in view of the different (3- versus 10-day) retention abilities involved, as well as a high degree of relationship between the two retention abilities. Test questions covered principles, facts, and applications, and were selected by an item analysis procedure from a larger population of items. Scores on both tests showed a satisfactory range of variability and were distributed normally. Since they were intended as power tests, no time limit was imposed.

Three types of introductory passages were used in this experiment, each about 500 words in length. The *comparative* organizer pointed out explicitly the principal similarities and differences between Buddhist and Christian doctrines. This comparison was presented at a much higher level of abstraction, generality, and inclusiveness than the Buddhism passage itself, and was deliberately designed to increase discriminability between the two sets of concepts. The *expository* organizer, on the other hand, merely presented the principal Buddhist doctrines at a high level of abstraction, generality, and inclusiveness, without making any reference whatsoever to Christianity. It was not explicitly designed to increase discriminability between the two religions, but merely to provide some general ideational scaffolding for the detailed Buddhist material. The *historical* introduction was intended solely as a control treatment. It contained interesting historical and human interest material about Buddha and Buddhism, but neither provided any ideational scaffolding nor attempted to compare Buddhism and Christianity. No information was included in any of the introductory passages that could constitute a direct advantage in answering questions on the Buddhism test.

It was methodologically important to provide an historical introduction for the control group in order that any obtained differences in retention outcomes between experimental (comparative) and control (historical) groups could be attributed to the particular nature of the comparative organizer (i.e., to its enhancing effects on discriminability) rather than to its presence per se. The purpose of exposing another group of subjects to an expository introduction was to determine whether the comparative organizer could increase the discriminability of the learning passage over and above that which could be attributed to the influence of a simple organizer (i.e., to the mere provision of advance ideational scaffolding).

A 36-item multiple-choice test on Christianity was used to measure variability among our subjects with respect to the stability and clarity of those existing (Christianity) concepts within cognitive structure which were analogous to the Buddhism learning material, and hence potentially interfering. Test items were noncontroversial, dealing with Old and New Testaments, church history, and denominational beliefs and differences. The scores on this test were normally distributed and the corrected split-half reliability was .84.

Scores on the verbal portion of the SCAT were available for 65 subjects. This test had been previously administered as part of a battery of entrance examinations for incoming university freshmen.

## *Procedure*

On the first day of experimentation, all subjects took the Christianity test and then (after assignment to a treatment group) studied one of the three kinds of introductions for 8 minutes. Membership in a treatment group (comparative, expository, or

historical) was determined by random assignment. The population of each of the three treatment groups was also stratified by sex so that the proportion of men to women subjects would be the same in each group. This was necessary because of the women's significantly higher verbal SCAT scores. It was possible to administer all three treatments simultaneously because they consisted of identical appearing introductory passages (with identical sets of directions), differing only in content.

To equalize the possible effects of prior extended exposure to the Buddhism material, those few subjects (5% of the total) who had taken a course in comparative religion were equally distributed in random fashion, among the three treatment groups. It had also been assumed that random assignment of subjects would render the different treatment groups equivalent with respect to such factors as learning ability and knowledge of Christianity. This assumption was confirmed empirically by the finding that differences between these groups on SCAT and Christianity test scores were negligible.

In order to control for the effects of different instructor, situational, and classroom climate variables in the six sections, students *within each* section were equally divided among the three treatment groups. Since analysis of the data showed that homogeneity of variance prevailed, both on an intersectional as well as on an intergroup basis, for both the Christianity test scores and the two sets of Buddhism test scores, it was considered justifiable to treat the scores of the three treatment groups on each of these instruments as comparable random samples from the same population.

Two days after studying their designated introductions, all subjects read and studied the same Buddhism passage for 35 minutes. One form of the Buddhism test was administered to all subjects 3 days later, and an equivalent form of the same test was administered one week after the first test, or 10 days after the learning passage.

A special randomly assigned control group of subjects was constituted out of the six sections, along the same lines described above for the three treatment groups. The procedure followed with this special group was identical with that used for the comparative group except that subjects studied a 1,800-word passage on the endocrinology of human pubescence instead of the Buddhism passage. The purpose of this procedure was to ascertain to what extent mere knowledge of the comparative organizer (without any exposure to the Buddhism learning passage itself) could increase scores on the Buddhism tests beyond chance expectancy.

## Results and Discussion

Comparison of corresponding 3- and 10-day means of total treatment groups shows that retention loss during this interval was relatively slight. The loss was greatest in the historical group ($p = .05$), least in the expository group, and intermediate in the comparative group ($p < .10$). This low degree of retention loss is probably attributable both to the negatively accelerated shape of most retention curves, particularly in the case of meaningful material, and to the "rehearsal effect" of the 3-day Buddhism test on the subsequent test of retention. The retention scores of our subjects over the two intervals were highly correlated ($r = .79$).

### *Effect of Organizers on Retention*

On a 3-day basis, only the comparative organizer was effective in facilitating retention of the Buddhism material. Table 1 shows that the mean retention score of the total comparative group was significantly superior to both that of the historical group ($p = .05$) and that of the expository group ($p = .05$). However, practically all of this obtained difference between the comparative group, on the one hand, and the expository and historical groups, on the other, was derived from the below-median subgroups on the Christianity test. Within these below-median subgroups the differences between the means in question were much greater than the corresponding differences between total treatment groups, and their level of significance was also correspondingly higher ($p = .02$, $p = .005$).

It is apparent, therefore, that al-

though provision of ideational scaffolding in the form of an expository organizer did not enhance retention of the Buddhism passage over a 3-day interval, the combined scaffolding and explicit discriminability effects induced by the comparative organizer did significantly improve retention. The short-term retention loss in the control group was evidently small enough to preclude the possibility of a significant difference in retention attributable to the facilitating effects of a simple expository organizer. It is true that an expository type of organizer significantly increased 3-day retention in an earlier experiment (Ausubel, 1960); but the learning passage used then was more unfamiliar to subjects, and the latter also had the benefit of studying the organizer on two separate occasions.

As hypothesized, subjects with relatively superior knowledge of Christianity derived considerably less benefit from the comparative organizer than did subjects whose knowledge of Christianity was less impressive. This finding was consistent with the self-evident proposition that if discriminability of a learning passage is already high because of endogenous factors within cognitive structure (i.e., because of the clarity and stability of related established knowledge), less scope exists for the potentially facilitating influence of exogenously manipulated factors (i.e., organizers) designed to promote discriminability.

On a 10-day basis, both the comparative and expository total groups were significantly superior to the historical total group in retaining the Buddhism material ($p = .02$, $p = .05$), but the difference between the comparative and expository groups was negligible (see Table 1). As was true of the 3-day scores, however, most of the difference between organizer and control groups was derived from the below-median subgroups on the Christianity test. When the means of just these below-median subgroups were compared, the significance level of the comparative-historical and the expository-historical differences was enhanced ($p = .0025$, $p < .05$), and the difference between the comparative and expository groups was significant at the .10 level.

In comparing the 10- to the 3-day retention data, it appears first, that only over the longer time interval was the natural retention loss sufficiently great to provide scope for the limited facilitating influence of the scaffolding effects available from the expository organizer. Second, although the comparative organizer was not significantly more effective than the expository organizer over the longer interval when the results of all subjects in these groups were considered, there was a suggestive trend in this direction among the below-median subjects.

One explanation of the relatively small difference in 10-day retention attributable to the influence of explicit comparison per se, is the possibility that confronting the learner in advance with the major principles of Buddhism in a detail-free context (the expository organizer) *implicitly* increased the discriminability of his Buddhism ideas by enabling him to make his own comparisons with Christianity. Another plausible explanation is that by enhancing retention generally, the rehearsal effect induced by the 3-day Buddhism test had a leveling influence on the relative degree of facilitation that might have been expected from the two kinds of organizers. This interpretation is supported by the erosion of the significant retention difference between the comparative and expository

TABLE 1
MEAN RETENTION TEST SCORES OF EXPERIMENTAL AND CONTROL GROUPS ON BUDDHISM PASSAGE

| Knowledge of Christianity | Treatment group | | |
|---|---|---|---|
| | Comparative organizer | Expository organizer | Control (historical) |
| 3-day retention: | | | |
| Above-median | 23.50 | 22.50 | 23.42 |
| Below-median | 20.50 | 17.32 | 16.52 |
| Total | 21.83 | 19.91 | 19.97 |
| 10-day retention: | | | |
| Above-median | 21.79 | 22.27 | 20.87 |
| Below-median | 19.21 | 17.02 | 14.40 |
| Total | 20.31 | 19.65 | 17.63 |

Note.—Chance score on the multiple-choice test of 45 items is 9.0.

groups from the third to the tenth day.

The tendency noted above in the 3-day results—for only the below- as opposed to the above-median subgroups on the Christianity test to derive appreciable benefit from the organizers—also appeared in the 10-day data. Thus, in the learning and retention of unfamiliar ideational material that is relatable to established concepts in the learner's cognitive structure, both comparative and expository organizers appear to be effective only in those instances where existing (endogenously determined) discriminability between the two sets of ideas is inadequate as a consequence of the instability or ambiguity of the established concepts.

The special control group which only studied the comparative organizer (without any exposure to the Buddhism passage itself) made a mean score of 13.20 on the 3-day Buddhism test and a mean score of 13.45 on the 10-day test. These scores were significantly greater than the scores of a comparable naive group (which took the Buddhism tests without being exposed to either organizer or learning passage), but were substantially below those of the historical and two organizer groups.

It is quite unlikely, however, that the organizers per se directly furnished pertinent information enabling subjects to obtain higher scores on the Buddhism tests. In the first place, a deliberate effort was made to avoid providing such information in the two organizers. Second, if the organizers themselves had furnished useful information in answering test items, it would be difficult to explain why subjects in the expository group did not make higher scores than subjects in the historical group on the 3-day Buddhism test. Much more credible, therefore, is the explanation that exposure to the comparative organizer merely increased the general sophisticaton of subjects in the special control group about Buddhist concepts, and thereby helped them to eliminate misleads in the multiple-choice test questions. Thus, they were able to obtain better than chance scores without studying the learning passage itself. But when the learning passage was available (i.e., in the historical and two organizer groups), subjects were neither benefited by the general sophistication they obtained from an ideational organizer, nor handicapped by not possessing such sophistication.

### *Effect of Knowledge of Christianity on Retention*

The data clearly support the hypothesis that the discriminability of the Buddhism learning material varies as a function of the clarity and stability of the established concepts to which it is related (i.e., Christianity), and hence that Buddhism retention scores are positively correlated with knowledge of Christianity. Table 1 shows that within each treatment

TABLE 2
CORRELATIONS OF BUDDHISM RETENTION SCORES AND OTHER MEASURES

| Other measures | Buddhism retention scores | | | | | |
|---|---|---|---|---|---|---|
| | Comparative group | | Expository group | | Historical group | |
| | 3-day | 10-day | 3-day | 10-day | 3-day | 10-day |
| Christianity scores | .37* | .21 | .57** | .42** | .55** | .56** |
| Verbal ability scores | .62** | .60** | .75** | .79** | .58** | .52* |
| Christianity scores with verbal ability eliminated | .06 | −.07 | .40** | .14 | .40** | .43** |

* Significant at .05 level.
** Significant at .01 level.

group the mean retention score of the above-median group was significantly greater than the mean retention score of the below-median group. For the historical, expository, and comparative groups, these differences were significant at the .0005, .001, and .02 levels, respectively, on the 3-day Buddhism test, and at the .0005, .0005, and .10 levels, respectively, on the 10-day Buddhism test. The same trends are shown by the positive correlations between scores on the Buddhism and Christianity tests (see Table 2).

It is evident, therefore, that the clearer and more stable subjects' knowledge of Christianty was, the more discriminable this knowledge was from Buddhism concepts, and hence the higher the Buddhism retention scores were. And conversely, the less clear and more unstable subjects' knowledge of Christianity was, the less discriminable it was from Buddhism concepts, and the lower the Buddhism retention scores were.

The data also confirm the hypothesis that organizers (by virtue of their leveling effect on the endogenous discriminability advantage inherent in a clear and stable knowledge of Christianity) reduce the relationship between Christianity knowledge and Buddhism retention scores in proportion to their facilitating effect on retention. The leveling effect of the organizers (particularly the comparative) can be seen in Table 1 by comparing the relative magnitude of the differences between above-median and below-median subgroups in the various treatment groups. For historical, expository, and comparative groups the respective $t$ values of these differences were 4.86, 3.43, and 2.24 for the 3-day Buddhism scores, and 4.72, 3.99, and 1.59 for the 10-day Buddhism scores. Correlations between the Buddhism and Christianity scores show the same trend (see Table 2). The correlation between the Christianity and Buddhism scores was higher in the historical than in the comparative group for both the 3-day ($p = .12$) and 10-day ($p = .04$) retention intervals. The almost identical correlations, on the other hand, between Christianity and 3-day Buddhism scores in the expository and historical groups, reflected the ineffectiveness of the expository organizer in facilitating retention over the 3-day interval.

Because of the positive correlation of .44 ($p < .01$) between Christianity and SCAT scores, however, it was necessary to check the alternative hypothesis that the significantly higher retention scores of subjects in the above-median groups reflect superior verbal ability rather than superior knowledge of Christianity per se (endogenous discriminability). But since the availability of only 65 SCAT scores made it impossible to match for verbal ability those subjects within each treatment group who were above

and below the median score on the Christianity test, it was necessary to resort to partial correlation.

The partial correlations between the Buddhism and Christianity scores, with the effect of verbal ability eliminated, are shown for all three groups in Table 2. For the historical group the partial correlations of .40 and .43 (for the 3- and 10-day scores, respectively), although lower than the corresponding simple correlations of .55 and .56 were still significant at the .01 level. Three of the four other partial correlations, however, were negligible and nonsignificant, indicating that the corresponding simple correlations largely reflected the positive relationship between verbal ability and Buddhism retention scores. The intrinsic residual relationship between knowledge of Christianity and retention of the Buddhism material, after the effect of verbal ability was eliminated, was actually close to zero in these instances, because of the leveling influence of the organizers on the endogenous discriminability advantage conferred by superior knowledge of Christianity. The partial correlation between the Christianity and 3-day Buddhism scores remained significant at the .01 level in the expository group, in as much as the expository organizer did not facilitate retention over the 3-day interval.

Intergroup comparison of the partial correlations in Table 2 shows even more clearly than does corresponding comparison of the simple correlations between Buddhism and Christianity scores, that the organizers (in proportion to their facilitating effect on retention) reduced the significant relationship between knowledge of Christianity and retention of the Buddhism material. With the influence of verbal ability eliminated, the partial correlation between Christianity and 3-day Buddhism scores in both the historical and expository groups was significantly higher than in the comparative group ($p = .05$, $p = .05$). The partial correlation between Christianity and 10-day Buddhism scores in the historical group was higher than the corresponding partial correlations in the expository ($p = .08$) and comparative ($p = .01$) groups. In effect then, by being provided with a given type of discriminability aid (organizer), subjects possessing relatively little knowledge of Christianity (and hence little endogenous discriminability) were placed on approximately the same footing with respect to the discriminability variable and its effect on retention, as subjects possessing greater knowledge of Christianity (and greater endogenous discriminability).

Table 2 also shows that the correlation between verbal ability and Buddhism retention scores in the historical group (where the relationship was not influenced by interaction with organizer effects) was approximately the same as that between Christianity and Buddhism scores. But unlike the latter relationship, which was attenuated by interaction with the organizers, the correlation between verbal ability and Buddhism scores was slightly higher in the comparative than in the historical group, and suggestively but not significantly higher in the expository than in the historical group ($p = .17$, $p = .07$). At the very least, therefore, it is definite that neither organizer detracted from the strong positive relationship between verbal ability and the retention of the Buddhism material; and there was a suggestive tendency for verbal ability to have an even greater impact on the retention of the Buddhism passage

when subjects were given the benefit of an expository organizer.

## Summary and Conclusions

This experiment was concerned with the role of endogenous and externally manipulated discriminability in the learning and retention of unfamiliar ideational material (Buddhism) that was relatable to previously learned concepts (Christianity) already established in cognitive structure.

The learning task consisted of a 2,500-word passage on the principles of Buddhism. Two days before studying this learning passage, one experimental group studied a 500-word *comparative* organizer explicitly comparing the major ideas of Buddhism and Christianity. Another experimental group studied an *expository* organizer which made no reference to Christianity. A control group studied an *historical* introduction dealing with the history rather than with the ideas of Buddhism. Retention of the Buddhism material was tested 3 and 10 days after the learning session by means of equivalent forms of a multiple-choice test. For purposes of analysis, subjects in the various treatment groups (university undergraduates) were divided into above- and below-median subgroups in terms of their scores on an objective test of Christianity.

Subjects with greater knowledge of Christianity made significantly higher scores on the Buddhism retention test than did subjects with less knowledge of Christianity. This significantly positive relationship between Christianity and Buddhism test scores held up even when the effect of verbal ability was eliminated. Hence the data support the hypothesis that the learning and retention of unfamiliar verbal material varies positively with its discriminability from related, previous learned concepts established in cognitive structure, and that this endogenously determined discriminability is a function of the clarity and stability of the latter concepts.

On a 3-day basis only the comparative organizer was significantly effective in facilitating the retention of the Buddhism material, but over a 10-day interval both comparative and expository organizers were significantly effective. These facilitating effects of the organizers on retention outcomes, however, only applied to subjects in the below-median subgroups on the Christianity test. Thus in the learning and retention of unfamiliar ideational material that is relatable to established concepts in the learner's cognitive structure, both comparative and expository organizers appear to be effective only in those instances where existing discriminability between the two sets of ideas is inadequate as a consequence of the instability or ambiguity of established concepts.

## REFERENCES

Ausubel, D. P. The use of advance organizers in the learning and retention of meaningful verbal material. *J. educ. Psychol.*, 1960, **51**, 267–272.

Ausubel, D. P., & Blake, E., Jr. Proactive inhibition in the forgetting of meaningful school material. *J. educ. Res.*, 1958, **52**, 145–149.

(Received May 8, 1961)

## ON ADOLESCENCE

*By Peter Blos.* A psychoanalytic interpretation of adolescent development. Dr. Blos discusses the intrinsic conflicts, liabilities, tasks, and resolutions typical for each phase of adolescent development, using the concept of the self to describe the normal progressive personality consolidation achieved during these years. A valuable book for psychologists, psychoanalysts, and educators. *$5.00.*

## CLINICAL PROCESS

*By E. Kuno Beller.* This important book offers—for the first time—a systematic guide to recording clinical data as well as an original method for evaluating that data. Dr. Beller proposes a system for organizing clinical data in the treatment of personality disorders in children and their parents. He also presents a quantitative method for analyzing and evaluating clinical records in practice and research. *$10.00.*

**THE FREE PRESS OF GLENCOE, INC.**

A DIVISION OF THE CROWELL-COLLIER PUBLISHING COMPANY

60 Fifth Avenue. New York 11, N. Y.

# EDUCATIONAL PSYCHOLOGY

**The Principles of Learning in Teaching**

by

ARDEN N. FRANDSEN, Utah State University

*McGraw-Hill Psychology and Human Development in Education Series*

640 pages $7.50

The purpose of this text is to "aid teachers in meeting the challenge of modern education: the provision of better education for children and youth." It focuses the scientific knowledge we have accumulated both about the individual characteristics of children and youth and about how they learn on two ways of attaining high levels of accomplishment: (1) providing for each individual the kind of education from which he can profit most, and (2) improving the efficiency of learning of all students.

**McGraw-Hill Book Company, Inc.**
**330 West 42nd St.,**
**N.Y. 36, N.Y.**